Office of Technology Assessment

Congressional Board of the 97th Congress

TED STEVENS, *Alaska, Chairman*

MORRIS K. UDALL, *Arizona, Vice Chairman*

Senate

ORRIN G. HATCH
Utah

CHARLES McC. MATHIAS, JR.
Maryland

EDWARD M. KENNEDY
Massachusetts

ERNEST F. HOLLINGS
South Carolina

HOWARD W. CANNON
Nevada

House

GEORGE E. BROWN, JR.
California

JOHN D. DINGELL
Michigan

LARRY WINN, JR.
Kansas

CLARENCE E. MILLER
Ohio

COOPER EVANS
Iowa

JOHN H. GIBBONS
(Nonvoting)

Advisory Council

CHARLES N. KIMBALL, *Chairman*
Midwest Research Institute

EARL BEISTLINE
University of Alaska

CHARLES A. BOWSHER
General Accounting Office

CLAIRE T. DEDRICK
California Land Commission

JAMES C. FLETCHER
University of Pittsburgh

S. DAVID FREEMAN
Tennessee Valley Authority

GILBERT GUDE
Congressional Research Service

CARL N. HODGES
University of Arizona

RACHEL McCULLOCH
University of Wisconsin

WILLIAM J. PERRY
Hambrecht & Quist

DAVID S. POTTER
General Motors Corp.

LEWIS THOMAS
*Memorial Sloan-Kettering
Cancer Center*

Director

JOHN H. GIBBONS

This is an OTA Technical Memorandum that has neither been reviewed nor approved by the Technology Assessment Board.

BEST AVAILABLE COPY

MEDLARS AND HEALTH INFORMATION POLICY

A TECHNICAL MEMORANDUM

SEPTEMBER 1982

Technical Memorandum are issued by OTA on specific subjects analyzed in recent OTA reports or on projects presently in process at OTA. They are issued at the request of Members of Congress who are engaged in committee legislative actions which are expected to be resolved before OTA completes its assessment.

CONGRESS OF THE UNITED STATES
Office of Technology Assessment
Washington, D.C. 20510

Library of Congress Catalog Card Number 82-600639

For sale by the Superintendent of Documents,
U.S. Government Printing Office, Washington, D.C. 20402

Preface

Accessible information is fundamental to the success of activities to improve the health of the population. With the increasing growth of the health literature, however, it is more and more difficult to locate needed information.

Both governmental and private sector organizations create computerized health-related data bases and provide access to them. Government-sponsored health-related bibliographic information is predominantly created and disseminated by the National Library of Medicine's (NLM) computerized system, MEDLARS. One of NLM's legislative mandates is to "aid the dissemination and exchange of scientific and other information important to the progress of medicine and the public health" (Public Law 84-941).

The House Interstate and Foreign Commerce Committee requested OTA to examine MEDLARS' performance as part of a larger OTA study, *Strategies for Medical Technology Assessment.* The Senate Labor and Human Resources Committee then requested OTA to explore the relationship between NLM and the private sector in creating and disseminating health-related information by means of computerized bibliographic retrieval systems.

This technical memorandum analyzes the arguments for and against NLM's creation and dissemination of health-related bibliographic information. It provides information designed to help Congress in decisions regarding the appropriate mix of NLM and private sector activities that might serve the public interest most efficiently.

A principal finding is that in most respects MEDLARS is an efficient system for disseminating health-related bibliographic information. In regard to the relationship of NLM and the private information sector, OTA's main findings are: 1) that there are insufficient empirical data to decide, on purely technical grounds, the most efficient and effective configuration of public and private bibliographic activities; and 2) that rapid developments in the computer and communications fields may, in the not too distant future, profoundly alter the effects of current decisions.

This study was guided by an advisory panel chaired by Dr. Robert Hayes. In addition, a large number of academics in the health and information fields, practicing health professionals and information specialists, and public and private information providers were consulted. We are grateful for their contributions.

John H. Gibbons

JOHN H. GIBBONS
Director

Advisory Panel for Strategies for Medical Technology Assessment

Lester Breslow, *Panel Chairman*
School of Public Health, University of California, Los Angeles

Morris Collen
Director of Technology Assessment
Kaiser-Permanente Medical Group

Richard Cooper
Williams & Connolly, Inc.

D. V. d'Arbeloff
Chairman and Chief Executive Officer
Millipore Corp.

Harvey Fineberg
Harvard School of Public Health

Jerome D. Frank
The Henry Phipps Psychiatric Clinic
The Johns Hopkins Hospital

William Goffman
School of Library Science
Case Western Reserve University

Leon Greene
Vice President
New Product Technology
Smith, Kline & French Laboratories, Inc.

David B. Horner
Orthopedic Surgeon

Stanley B. Jones
Vice President
Blue Cross/Blue Shield Association

F. Wilfrid Lancaster
Graduate School of Library and Information Science
University of Illinois

Louise B. Russell
Senior Fellow
The Brookings Institution

Herbert Semmel
President
Consumer Coalition for Health

Robert M. Veatch
Kennedy Institute of Ethics
Georgetown University

Richard W. Vilter
American College of Physicians
College of Medicine
University of Cincinnati

Kenneth E. Warner
School of Public Health
University of Michigan

Richard N. Watkins
Staff Physician
Group Health Cooperative

Carol Weiss
Graduate School of Education
Harvard University

Kerr L. White
Deputy Director for Health Sciences
Rockefeller Foundation

OTA Project Staff—MEDLARS and Health Information Policy

H. David Banta, *Assistant Director, OTA*
Health and Life Sciences Division

Clyde J. Behney, *Health Program Manager*

Bryan Luce, *Project Director for Strategies for Medical Technology Assessment*
Gloria Ruby, *Study Director*
Dale A. Carlson, *Analyst*
John C. Langenbrunner, *Analyst*
Kerry Britten Kemp, *Editor and Writer*
Virginia Cwalina, *Administrative Assistant*
Mary E. Harvey, *Secretary*
Pamela J. Simerly, *Secretary*
Nancy L. Kenney, *Secretary*[1]
Lorraine G. Ferris, *Secretary*[2]
Michael P. Hughes*

Principal Contractors

Jose-Marie Griffiths, *King Research Associates*

Tefko Saracevic, *Case Western Reserve University*

Richard Solomon, *Massachusetts Institute of Technology*

Ithiel de Sola Pool, *Massachusetts Institute of Technology*

Patricia Woolf, *Princeton University*

OTA Publishing Staff

John C. Holmes, *Publishing Officer*

John Bergling Kathie S. Boss Debra M. Datcher Joe Henson

---

[1] Until September 1981.
[2] Until January 1982.
*Contract personnel.

Advisory Panel for MEDLARS and Health Information Policy

Robert Hayes, *Chairperson*
Dean, Graduate School of Library and Information Science,
University of California, Los Angeles

Robert L. Chartrand, *ex officio*
Senior Specialist in Information Policy and Technology
Congressional Research Service

Robert Cheshier
Director
Cleveland Health Sciences Library

Don Detmer
University of Wisconsin Medical School

Donald Dunn
Department of Engineering Economic Systems
Stanford University

Rashi Fein, *ex officio*
Center for Community Health and Medical Care
Harvard Medical School

Morton David Goldberg
Schwab, Goldberg, Price & Dannay

William N. Hubbard, Jr.
President
The Upjohn Co.

Lee Hyde
Assistant Director
Kingsport Family Practice Center
Kingsport, Tenn.

Edward Kennedy
President
BIOSIS

F. Wilfrid Lancaster
University of Illinois

Davis McCarn
Director
Computerized Bibliographic Services
The H.W. Wilson Co.

Judith Messerle
Medical Information Center
St. Joseph Hospital
Alton, Ill.

Christopher A. Meyer, ex officio
Senior Attorney Office of General Counsel
U.S. Copyright Office

Barbara Quint
Head, Reference Services
Rand Corp.

Roger Summit
President
DIALOG Information Services, Inc.

Robert Wedgeworth
Executive Director
American Library Association

Carol Weiss
Graduate School of Education
Harvard University

Gloria Werner
Biomedical Library
University of California, Los Angeles

Kerr L. White
Deputy Director of Health Sciences
Rockefeller Foundation

Contents

Chapter *Page*

1. INTRODUCTION AND BACKGROUND 3
  History, Objectives, and Scope of the Study 3
  Background for the Study 4
  Organization of the Report 6

2. MEDLARS 11
  Introduction 11
  Development of MEDLARS 11
  Current Status of MEDLARS 13
    Data Bases 13
    Leasing 16
    On-Line Services Users and Uses 16
  The Future of MEDLARS 23
    Current Standing 23

3. MEDLARS: SYSTEM-RELATED ISSUES 27
  Introduction 27
  Subject Coverage 27
  Literature Coverage 29
  Methodological Merit 30
  Discussion 31

4. PRIVATE SECTOR HEALTH INFORMATION SYSTEMS 35
  Introduction 35
  Historical Origins 35
  Data Bases Relevant to Health Sciences 36
  BIOSIS PREVIEWS 36
  EXCERPTA MEDICA 37
  IRL LIFE SCIENCES COLLECTION 38
  SCISEARCH 38
  Commercial Search Services 39
    Bibliographic Retrieval Services, Latham, N.Y. 40
    DIALOG Information Services, Inc., Palo Alto, Calif. 41
    System Development Corp., Santa Monica, Calif. 42

5. PUBLIC AND PRIVATE INFORMATION SECTORS: ELEMENTS OF CURRENT DOMESTIC POLICY ISSUES 47
  Introduction 47
  Definitions 48
  Historical Perspective 48
  The Appropriate Role of Government in Information Activities: Major Underlying Considerations 50
    Introduction 50
    Allocation of Resources 51
    Effects of Government Involvement in Information Activities on the Private Information Sector 52
    Pricing of Information Products and Services 54

6. MEDLARS AND PRIVATE HEALTH INFORMATION SYSTEMS: DISCUSSION OF DOMESTIC POLICY ISSUES 65
  Range of NLM's Computerized Products and Services 66
    Data Bases 66
    On-Line Services 70

Chapter *Page*
Pricing Issues 72
Fees for Leasing Data Base Tapes 72
On-Line Service Charges 74
Differential Pricing 76
Research and Development 78
Final Comments 78

Appendixes

A. The National Library of Medicine: Organization and Activities 83
B. U.S. Information Policies 99
C. MEDLARS Evaluations: A Review of the Literature 104
D. Development of Computerized Biomedical Bibliographic Retrieval Systems 109
E. MEDLINE: Technical Processes 111
F. Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science: Principles and Recommendations 114
G. AGRICOLA and ERIC 116
H. Future Information Technologies: Implications for Biomedical Retrieval Systems 120
I. Issues in International Access to MEDLARS 128
J. Health Program Advisory Committee and Acknowledgments 138
K. Acronyms, Abbreviations, and Glossary 140

References 145

List of Tables

Table No. *Page*
1. NLM Data Bases, Fiscal Year 1981 13
2. NLM Data Bases and Updates, Fiscal Year 1981 14
3. MEDLARS: Usage of NLM Data Bases During September 1981 15
4. MEDLARS: Number of Domestic Institutions On-Line by Region and Type, September 1981 17
5. Summary Comparison of the Functional Capabilities of MEDLARS II and MEDLARS III 24
6. Excerpta Medica Abstract Journals 37
7. IRL Abstracts 38
8. On-Line Connect Hour Rates for Selected Health-Related Data Bases, Spring 1982 40
9. BRS: Selected Health-Related Data Bases Available On-Line, March 1982 41
10. DIALOG: Selected Health-Related Data Bases Available On-Line, April 1982 43
11. SDC: Selected Health-Related Data Bases Available On-Line, March 1982 44
12. Selected Data Bases Distributed by NTIS, December 1981 73
A-1. NLM Appropriations, 1970-82 87
A-2. NLM Staff, Fiscal Years 1975-81 87
A-3. Medical Library Assistance Act Authorizations and Appropriations, Fiscal Years 1975-82 89
A-4. Medical Library Assistance Act: Distribution of Funds Among Grant Programs, Fiscal Years 1980-82 90
A-5. Medical Library Assistance Act: Resource Grants, Fiscal Years 1971-82 92
A-6. Medical Library Assistance Act: Training Grants, Fiscal Years 1971-82 93
A-7. Medical Library Assistance Act: Research Grants, Fiscal Years 1971-82 95
A-8. Medical Library Assistance Act: Special Scientific Project Grants, Fiscal Years 1971-82 . 95
A-9. Medical Library Assistance Act: Publication Grants, Fiscal Years 1971-82 96
C-1. MEDLINE's Coverage of Selected Topics 108
G-1. ERIC Clearinghouses and Operating Organizations, 1981 118
I-1. Foreign Centers' Access to MEDLARS, March 1982 128

List of Figures

List of Figures *Page*

1. MEDLARS: Number of Domestic and Foreign Institutions On-Line, 1971-81 18
2. MEDLARS: Percent of Searches and Connect Hours by User Categories, Fiscal Year 1981 19
3. MEDLARS: Use of MEDLINE and HEALTH On-Line Data Bases by Connect Hours by User Categories, Fiscal Year 1981 19
4. MEDLARS: Number of On-Line Searches Performed, 1972-81 20

A-1. National Library of Medicine 84

A-2. Representative NLM Services, 1970-81 88

A-3. Medical Library Assistance Act Appropriations and Regional Medical Library Program Budget Current and Constant Dollars, Fiscal Years 1971-81 90

A-4. Distribution of Resource Grants, Fiscal Years 1971-78 92

Chapter 1

Introduction and Background

Chapter 1

Introduction and Background

The significance of information to society is becoming increasingly apparent with advances in technology. Most nations today realize that they must be able to obtain and efficiently process information vital to national life. They must especially be able to do this with respect to health information. The transfer of scientific information to researchers who require it in the conduct of their investigations and to practitioners of health care is essential for the health of the American people.

The National Library of Medicine (NLM, or the Library) is the Nation's—and in many respects the world's—principal resource for the collection, organization, and retrieval of scientific literature in the health and biomedical fields. Its efforts complement the Nation's investment in biomedical and other health-related research, and in medical education. NLM's purpose, as expressed in the National Library of Medicine Act (Public Law 84-941), is "to assist the advancement of medical and related sciences, and to aid the dissemination and exchange of scientific and other information important to the progress of medicine and to the public health."

NLM is a complex organization that performs diverse and far-reaching activities, extending well beyond what is customarily considered as traditional librarianship. (See app. A for a description of NLM, including its organization, appropriations and staffing, and intramural and extramural activities.) It has funded grant programs that have enhanced the collections of academic health center and community hospital libraries, promoted library consortia, and supported the research and development of computer applications to medicine as well as to information science. The Library has also supported the development of a national system of regional medical libraries, which links NLM and local libraries.

HISTORY, OBJECTIVES, AND SCOPE OF THE STUDY

Two congressional requests to OTA prompted this study. While expressing different concerns, the two are complementary. The first request was part of a larger request from the House Interstate and Foreign Commerce Committee for a study addressing the techniques and methods available for assessing medical technologies.* The committee specifically asked OTA to examine the performance of MEDLARS, particularly the performance of its major biomedical data base MEDLINE, in disseminating health-related bibliographic information.** The second request, from the Senate Labor and Human Resources Committee, resulted from interest in issues raised in an OTA staff paper on NLM and from hearings on reauthorization of the Medical Library Assistance Act held in April 1981. It asked that OTA explore the Government's role in the creation and the distribution of health-related information by means of computerized bibliographic retrieval systems.

Thus, this technical memorandum has two objectives. The major objective is to examine NLM's role in the creation and distribution of computerized health-related bibliographic information in light of the private sector's presence in this field and the public interest. NLM's role with respect to the transfer of bibliographic information is not a unique issue, but is part of a growing concern and discussion about the Government's role in the creation and the dissemination of all types and forms of information.*

---

*See OTA's study *Strategies for Medical Technology Assessment* (117).

**Since the inception of NLM's computerized system the terms MEDLARS and MEDLINE have had a number of definitions. The most recent definitions and those used throughout this technical memorandum follow. MEDLARS (Medical Literature Analysis and Retrieval System), NLM's computerized retrieval and technical processing system, is a complex IBM multiprocessing system that maintains data files, provides on-line retrieval services and produces computer-photocomposed publications. MEDLINE (MEDLARS on-line) is the largest and most extensively used of NLM's data bases.

*This issue is but one of many national and international policy issues. App. B provides a contextual setting for the specific infor-

(continued on next page)

The second objective, which stems from the House Interstate and Foreign Commerce Committee request, is to examine MEDLARS' effectiveness in disseminating bibliographic health-related information. Part of the response to that request is contained in the OTA staff paper "The National Library of Medicine" (116) and in the OTA report *Strategies for Medical Technology Assessment* (117). This technical memorandum analyzes three system relevant issues: 1) the subject content of the literature cited in the MEDLARS data bases, in particular MEDLINE; 2) the coverage of the nonserial literature in the MEDLARS data bases, in particular MEDLINE; and 3) the evaluation of the methodology and statistics used in the literature cited in the MEDLARS data bases. OTA was not requested to, and has not attempted to, analyze the effectiveness of private sector computerized bibliographic health-related information products and services. Nonetheless, the issue of assessing the adequacy of methodological design and statistical analysis in cited literature pertains to all health-related data bases.

BACKGROUND FOR THE STUDY

One of the mandated functions of NLM is the collection and preservation of health-related primary library materials, such as books, periodicals, prints, films, and recordings. Another mandated function is to organize the primary literature and publish and make available indexes, catalogs, and bibliographies—i.e., secondary literature—in order to locate relevant primary literature.

Although both the public and private sectors have a long history of providing bibliographic biomedical information products and services—i.e., secondary literature—the Library of the Office of the Army's Surgeon General, the forerunner of NLM, began the first printed index to the biomedical literature in 1879 with *Index Medicus*. The index, which primarily contains references to published journal articles, was prepared manually until 1964, when MEDLARS mechanized the processing and printing functions. The index records were put into machine-readable form, thereby expediting the production of the printed *Index Medicus* and making these records usable as a machine-readable data base. Thus, in 1964, NLM started the first large-scale, computer-based, retrospective search service available to the general public. In 1966, as a result of research in both the public and private sectors, Lockheed Information Systems (now DIALOG Information Services, Inc.), a commercial firm, made available the first on-line search service on a regular production basis.*

Today, the data tape used for the printing of *Index Medicus* and other NLM printed publications is also used as the source of data for the data base MEDLINE. Subsets of the MEDLINE data base are incorporated into some of the other data bases that are now available on-line at NLM. In addition to producing data bases and data base products, NLM now provides direct on-line machine searching of the contents of the data bases to the information and health communities. The Library is connected by a complex telecommunications network to more than 1,800 terminals in institutions in the United States and abroad. It also leases tapes of its data bases to two commercial U.S. firms which disseminate the information online from their computer to their customers. Organizations in a number of foreign countries also lease some of the data bases.

Worldwide, *Index Medicus* has been the primary means for access to medical information for more than 100 years. In the past 17 years, MEDLARS has likewise achieved an impressive

---

(continued from p. 3)

mation policy issues of interest by briefly reviewing the history, current standing, and future prospects of domestic information policy and describing the relationship between domestic and international information policy.

*With on-line access, a person at a computer terminal can carry on a dialog with the computer and direct it to locate information, retrieve it, and provide it either at the terminal or in printed form for mailing to the requestor.

reputation. By providing on-line access to its bibliographic data bases, MEDLARS has made NLM's resources more accessible operationally and geographically to almost all segments of the biomedical community. Furthermore, NLM has sought to maintain a coverage of the biomedical literature that includes subjects of current interest, and the Library now serves a diverse constituency including many health science disciplines outside of traditional medicine.

There are, however, two current issues that challenge NLM and MEDLARS. *One issue is how to establish a suitable equilibrium between MEDLARS and the changing needs of its users.* Although the scope of MEDLARS data bases is limited, the quantity of knowledge in traditional biomedical subjects is increasing. Furthermore, much current biomedical research is interdisciplinary, and the present concept of health is broad and information pertaining to health can be found in journals in fields such as law and economics that are historically not the Library's province. Moreover, health-related information is often found in a form, such as technical reports and speeches, that is not cited in MEDLINE.

The expansion of the biomedical literature base has been accompanied by rapid progress in the application of computer and communications technologies to information systems, and a tremendous increase in the volume and ease with which information can be accessed. In order to assist the selection process for the health professional who uses the information that is retrieved and to provide a modicum of quality control, the scientific merit of literary material requires more and more scrutiny. One suggestion is that data base producers such as NLM assume some of the responsibility for assessing the evaluative methodology and statistical analysis used in the documents cited in bibliographic data bases.

The problem of accommodating user needs and MEDLARS' limitations is compounded because of the lack of comprehensive data on MEDLARS users. Although there is information available on the location of the terminals with access to MEDLARS by type of institution, information on the ultimate users of the information is sparse and insufficient for defining specific segments of the user population. (System issues are discussed in ch. 3.)

The second and more pressing problem facing NLM is its role in the creation and distribution of computerized health-related bibliographic information through MEDLARS. Not only NLM but for-profit and not-for-profit organizations in the private sector create health-related bibliographic data bases; such organizations include BIOSIS (formerly BioSciences Information Service), Excerpta Medica, Information Retrieval, Ltd., and the Institute for Scientific Information. In addition, biomedical data bases produced by both the public and private sectors are vended by three commercial organizations—Bibliographic Retrieval Services, DIALOG Information Services, Inc., and System Development Corp. Nonetheless, health-related computerized bibliographic information is predominantly created and disseminated by NLM through MEDLARS. At issue is whether NLM's computerized bibliographic products and services and the products and services of the private sector substitute for or complement each other and whether NLM's leading portion in the biomedical information field is hindering the growth of the private information sector.

Some members of the information community, and some members of the commercial sector of the private information industry, have become increasingly concerned about NLM's dominant role in the field. Because the industry is heterogeneous and composed of a variety of firms, it does not have one position regarding all of NLM's computerized bibliographic activities. Individual firms have particular opinions about NLM's activities depending on their perspectives. Overall, the industry's concerns are with respect to NLM's preparation of computerized bibliographic health-related data bases, NLM's charges to commercial information services and foreign centers for leasing the data tapes, NLM's provision of direct on-line access to its data bases, and NLM's pricing of direct on-line access to the data bases. NLM's pricing of access appears to be the issue of paramount concern at this time. For the most part, prices are significantly lower than those charged by the private sector for access to health-related bibliographic information. (These issues are discussed in ch. 6.)

Historically, public policy has held that NLM's position in the creation and dissemination of health-related bibliographic information is in the best interests of the public's health and well-being. Policy concerning the Library has been closely tied to biomedical research policy. The political environment in this regard appears to be changing. Although Government-sponsored biomedical research still appears to be of major interest, recent announcements from the Office of Management and Budget stress the importance of increased private participation in Government information activities and the need for "full cost recovery" (undefined) of Government-sponsored information products and services. The Library's role in providing computerized health-related bibliographic information may be reduced if such suggestions are implemented.

This OTA analysis focuses on current NLM issues as they relate to national information policy. These issues have direct importance for institutional and individual users of health-related bibliographic data bases, NLM, private sector producers of health-related data bases, and private information retrieval services (vendors). The issues are also significant for foreign data base producers and information retrieval services, both governmental and private, and for foreign institutional and individual users of health-related bibliographic data bases. Because the issues are similar to those concerning other Federal information activities, their resolution also has implications for Federal and non-Federal organizations that create and distribute Government-sponsored information, private sector information enterprises, foreign information organizations, and all users of Government-sponsored information.

This report's emphasis on current issues of necessity gives inadequate attention to the future effects of new and emerging technologies on biomedical bibliographic retrieval systems. Current issues result in part from technologies now in use. Evolving computer and communications technologies are still only in the early stages of development, and indications are that they will be much more powerful and varied in the future. *With expected dramatic changes in data base creation and access, some current issues may diminish in importance or disappear, and quite different ones may arise and require consideration.*

But it is not possible to know with any degree of certainty which technologies will be adopted and which issues will become significant. Social, political, and economic forces, as well as technological forces, are instrumental in determining the development and utilization of any innovation. Indeed, *the way decisions are made about current issues may affect which new and emerging technologies are implemented. This underscores the need for flexibility in public policy.*

There are indications that in the future there will be more distributive means of disseminating information thar are utilized at present.* Indeed, NLM administrators think that the technologies will be available in the next 5 to 7 years; they are making long-range plans in line with this thinking. Currently, NLM is experimenting with making one of its data bases available for distributive searching in 1 or 2 years. Thus, future technologies and their possible effects on biomedical bibliographic retrieval systems are discussed briefly in this study (see app. H).

ORGANIZATION OF THE REPORT

Chapter 2 describes the development, current status and future plans for MEDLARS' data bases and on-line services. System issues related to the effectiveness of MEDLARS in disseminating bibliographic health-related information are discussed in chapter 3. Chapter 4 is a brief description of private sector health-related data bases and commercial information services. The next chapter,

---

*A distributive data-processing system is one which uses multiple small computers to process all or portions of a data base. The small computers can be widely separated and may be linked by telecommunications lines to each other and to a large computer.

chapter 5, discusses the considerations underlying the current debate on the appropriate role of the Government in information transfer. Finally, chapter 6 analyzes the domestic and international implications of changing the range and pricing structure of MEDLARS computerized products and services. That chapter also considers the effect of new technologies on present issues. There are 11 appendixes included—both for reference and, in some cases, for expanding ideas and issues contained within the body of this report.

Chapter 2

MEDLARS

Chapter 2
MEDLARS

INTRODUCTION*

The dramatic rise in the number and types of biomedical primary publications, such as books, journals, and technical reports, often makes direct access to this literature difficult and confusing. Increasingly, secondary publications, including indexes, bibliographies, abstracts, and catalogs, serve as important elements in information transfer by directing users to primary sources. Development and use of computerized bibliographic information systems have further facilitated access to the growing body of primary health-related information publications.

The National Library of Medicine (NLM or the Library) has been in the forefront of this field and pioneered the "first large-scale library-based reference retrieval system" for health-related information (34). Since its inception, MEDLARS has become more sophisticated, accessible, and inclusive; and MEDLARS II, the Library's current system, is utilized more than any other system by health communities in the United States and abroad.

This chapter describes the development of MEDLARS, its current operations, and its projected future operations, providing information for the discussion of system issues in chapter 3 and for the analysis of public/private issues in chapter 6. As requested by Congress, the primary focus of the discussion is on MEDLINE, MEDLARS's major biomedical bibliographic data base.

DEVELOPMENT OF MEDLARS

The development of MEDLARS parallels the evolution in medical bibliography that began in the 1950's with the first attempt to apply computer technologies to information processing. (See app. D for a general description of the development of computerized biomedical bibliographic retrieval systems.) The system's first data base, MEDLINE, was a byproduct of the computerized production of the printed *Index Medicus.* One of the earliest accomplishments of NLM's predecessor, the Library of the Army's Surgeon General, is attributed to John Shaw Billings and Robert Fletcher, who together began the first monthly index of the world's periodical medical literature in 1879 (74).

The early history of *Index Medicus* was marked by severe financial deficits, frequent changes in sponsorship, and the commitment of a few dedicated individuals and organizations. The Army Medical Museum and Library, the successor to the Surgeon General's Library, struggled to keep the index afloat despite financial problems until 1899, when the publication failed. It was then revived for 3 years as *Bibliographia Medica* by a group of French physicians. In 1903, the Carnegie Foundation undertook its financial sponsorship, with Fletcher as chief editor. In 1927, *Index Medicus* was merged with a similar index produced by the American Medical Association, and its circulation quickly increased sixfold, though it continued to lose $25,000 to $50,000 annually. In 1960, NLM assumed responsibility for publishing *Index Medicus* monthly, and the American Medical Association began the annual *Cumulated Index Medicus (CIM).* In 1965, with the advent of computer applications to publishing, the American Medical Association turned *CIM* over to NLM (74).

Index Medicus continues to provide physicians and other health professionals the major access to biomedical literature worldwide (142). Published monthly, *Index Medicus* cites articles from 2,600 biomedical periodicals published in 36 languages. NLM endeavors to include references to published articles in periodicals "judged to be of greatest potential use to the international biomedical community" (35). In 1879, Billings' and Fletcher's index cited some 20,000 articles. One hundred years later, *Index Medicus* adds citations of about 220,000 articles annually, and about 6,000 biomedical institutions subscribe to it each year (36).

---

*Information in this chapter was obtained primarily from NLM staff and NLM publications.

By the late 1950's, NLM recognized that the exponential growth in biomedical literature required revision of the manual procedures for preparing *Index Medicus* and other publications. With the aid of a $73,800 grant from the Council on Library Resources in 1960, the Library developed a mechanized process that improved the method of preparing citations for publication.

In 1964, a true computerized system, MEDLARS, was put in place. The system was developed between August 1961 and December 1963, by the General Electric Corp., under contract to NLM. MEDLARS cost $3 million, some of which was supplied, at the recommendation of the National Advisory Heart Council, by the National Heart Institute. The new system improved the quality of the printed *Index Medicus*, enlarged its size, and decreased the time required for its printing by using the first computer-driven photocomposing device, Graphic Arts Composing Equipment (GRACE). The research and development of GRACE by the Photon Corp. was sponsored by NLM. The new technology represented a significant advance in the technology of computer typesetting (3). It also allowed for an increase in the number of articles that could be included in *Index Medicus* and in the number of subject terms assigned to each article. A module supporting the production of NLM's published catalog and its catalog cards was added in 1965. MEDLARS was used to produce all of NLM's publications from 1964 to 1975.

It was evident to NLM that the data base prepared for *Index Medicus* could be used for machine searching. In 1964, therefore, the Library began a batch computer operation. Users of MEDLARS would telephone, mail their requests, or make personal visits to the Library and other selected sites, and trained search analysts would access the system for the designated information. However, there was no direct interaction between the searchers and the computer, the process was expensive, and the time between the submission of a request and receipt of the resulting bibliography ranged from 3 to 6 weeks.

In 1968, NLM began planning a more advanced, *on-line* automated support system to create new files and allow on-line formulation of searches. The specification and test of the logic of the search was to be done on-line, but the actual searching were to be done later in the batch mode. Meanwhile, in cooperation with the System Development Corp. (SDC), NLM was experimenting with an on-line bibliographic retrieval system. This system, AIM-TWX (*Abridged Index Medicus* via the Teletypewriter Exchange Network), used an abridged data base consisting of 5 years of citations from the 100 most important English-language journals in clinical medicine. The system became operational in May 1970, and was so successful that by November of the same year over 80 institutions were using it. AIM-TWX ran on SDC's IBM 360/67 computer. Telecommunications enabling remote access were provided via the TWX network (then part of the American Telephone & Telegraph Co. but subsequently transferred to the Western Union Corp.).

But the new MEDLARS, MEDLARS II, took more than 6 years to be completed. The original contractor could not fulfill its obligations, and a contract was negotiated with SDC. As an intermediate solution, SDC merged the old MEDLARS I with the already successful on-line retrieval system AIM-TWX and put MEDLINE (MEDLARS On-line) service in place in October 1971.* Because the communications facilities provided via the TWX network were disproportionately high in cost, NLM contracted with Tymshare, Inc., in 1971, to provide data communications services for MEDLINE. The MEDLINE data base included citations from 1,000 to 1,200 journals for the current year and the previous 3 years. The completely new MEDLARS II, which finally became operational in January 1975, is the system now in place. The computer was successfully connected to TYMNET, a telecommunications network, in 1977.

---

*Currently, the term MEDLINE is mainly used to describe the data base and not the service.

CURRENT STATUS OF MEDLARS

NLM is continuously refining MEDLARS II to extend its capabilities. The system has evolved into a complex multiprocessing system that maintains data files, provides on-line retrieval services, and produces computer-photocomposed publications. The main computer facility is located in the NLM building in Bethesda, Md.; a second facility for processing searches is at the State University of New York (SUNY), Albany. In September 1980, NLM enlarged its computer capacity twofold with the transfer of an IBM 370/168 multiprocessor system from the National Institutes of Health's (NIH) Division of Computer Research and Technology.

Commercial firms and the Government cooperate in providing telecommunications to NLM's computerized data bases through a "complex arrangement of lease lines, public dial-up switched circuits, commercial telecommunications networks, and the U.S. Government-sponsored research network (ARPANET)" (134). TELENET and TYMNET are the principal commercial networks and carry 80 percent of the on-line traffic.

Data Bases

General Description

MEDLARS II contains almost 20 data bases. The machine-readable data tape that is the source of data for MEDLINE is used to produce *Index Medicus* and other publications including the *Abridged Index Medicus, Health Science Serials, National Library of Medicine Audiovisuals Catalog,* lists of citations in specialized biomedical fields (termed recurring bibliographies), and individual bibliographic searches considered to be of general interest. Table 1 lists the scope and date of initiation of NLM's current on-line data bases. Most of the data bases, like MEDLINE, are bibliographic, and contain references to the primary journal literature (monographs, serials, etc.); a few contain numeric or representational information.

Some of the data bases—MEDLINE, CATLINE, AVLINE, HISTLINE, SERLINE, and SDILINE—are created and maintained solely by NLM.

Table 1.—NLM Data Bases, Fiscal Year 1981

Data base and scope	Date initiated
AVLINE (*A*udio*V*isuals on-*Line*)—citations and abstracts to about 10,000 health science audio-visuals	February 1976
BIOETHICSLINE (*Bioethics* on-*Line*)—13,000 references to materials on bioethical topics	March 1978
CANCERLIT (*Cancer LI*terature)—285,000 references dealing with various aspects of cancer	July 1974
CANCERPROJ (*Cancer* research *PROJ*ects)—20,000 descriptions of ongoing cancer research projects	November 1975
CATLINE (*CAT*alog on-*LINE*)—300,000 references to books and serials cataloged at NLM since 1965	September 1973
CHEMLINE (*CHEM*ical dictionary on-*LINE*)—a file of some 1,000,000 names for chemical substances, representing 500,000 unique compounds	May 1974
CLINPROT (*CLIN*ical cancer *PROT*ocols)—summaries of clinical investigations of new anticancer agents and treatment techniques	February 1976
EPILEPSYLINE (*EPILEPSY* on-*LINE*)—contains about 25,000 references and abstracts to articles on epilepsy	January 1976
HEALTH (*HEALTH* Planning and Administration)—contains about 200,000 references to literature on health planning, organization, financing, management, manpower, and related subjects	November 1978
HISTLINE (*HIST*ory of Medicine on-*LINE*)—some 48,000 references to articles, monographs, symposia, and other publications dealing with the history of medicine and related sciences	August 1978
MEDLEARN—computer-assisted instruction program which teaches the novice user how to search the NLM on-line system	November 1976
MEDLINE (*MED*LARS on-*LINE*)—600,000 references to biomedical journal articles published in the current and 2 preceding years. An English abstract, if published with the article, is frequently included	October 1971
Backfiles that total some 2,700,000 references	
MED 66	November 1975
MED 69	December 1974
MED 72	May 1975
MED 75	December 1977
On-line	
MED 77	December 1979
MED 79	December 1981
MeSH Vocabulary File (*Me*dical *S*ubject *H*eadings Vocabulary File)—an on-line vocabulary file of the 14,000 medical subject headings that are used for subject cataloging, and also approximately 20,000 chemical records	April 1974
NAF (*N*ame *A*uthority *F*ile)—an authority list of 125,000 personal names, series names, corporate names and series decision records	September 1973
POPLINE (*POP*ulation Information on-*LINE*)—about 80,000 citations and abstracts to journal articles, monographs, and technical reports in the field of population	March 1980

Table 1.—NLM Data Bases, Fiscal Year 1981—continued

Data base and scope	Date initiated
RTECS (*R*egistry of *T*oxic *E*ffects of *C*hemical Substances)—an annual compilation of toxicity data for approximately 50,000 substances	July 1977
SDILINE (*S*elective *D*issemination of *I*nformation on-*LINE*)—references to the most current month of MEDLINE of approximately 20,000 citations	October 1972
SERLINE (*SER*ials on-*LINE*)—bibliographic information for about 38,000 serial titles, including all journals which are on order or cataloged for the NLM collection	November 1977
TDB (*T*oxicology *D*ata *B*ank)—contains chemical, pharmacological, and toxicological information and data on approximately 3,500 substances	July 1977
TOXLINE (*TOX*icology Information on-*LINE*)—a collection of about 600,000 references from the last 6 years on published human and animal toxicity studies, effects of environmental chemicals and pollutants and adverse drug reactions	April 1974
Contains 660,000 references to older materials	
TOXBACK 65	September 1980
TOXBACK 74—on-line	September 1980

SOURCE: National Library of Medicine.

MEDLEARN, MeSH and NAF, which are basically support files, are also products of the Library. The other data bases are supported or produced in collaboration with various institutions. BIOETHICSLINE is produced in cooperation with the Kennedy Institute of Ethics, Center for Bioethics at Georgetown University. CANCERLIT, CANCERPROJ, and CLINPROT are sponsored by the National Cancer Institute of NIH. HEALTH is produced in cooperation with the American Hospital Association and the Health Resources Administration. POPLINE is produced in cooperation with the Population Information Program of Johns Hopkins University and the Center for Population and Family Health at Columbia University. And RTECS is an on-line searchable version of a publication prepared by the National Institute for Occupational Safety and Health. CHEMLINE is created by NLM's Toxicology Information Program in collaboration with Chemical Abstracts Service. TOXLINE contains secondary information from Chemical Abstracts Service, BIOSIS (formerly Biosciences Information Service), the American Society of Hospital Pharmacists, the Environmental Protection Agency, Oak Ridge National Laboratory, the now defunct Smithsonian Science Information Exchange (SSIE) data base, the Hayes File on Pesticides, and a Toxicity Bibliography produced by NLM.*

MEDLINE

MEDLINE, with its related back files (MEDLINE from 1966 to 1977), is the largest data base in MEDLARS and is, by far, the most extensively used (see tables 2 and 3). The literature indexed for MEDLINE contains not only information on the science and practice of medicine and public health, but also on bioengineering, bioethics, and

Table 2.—NLM Data Bases and Updates, Fiscal Year 1981

Data base	Total records at data base[a]	Average number of records added per update	Frequency of updating
MEDLINE	638,374	22,000	Monthly
On-line and backfiles	3,631,463		
TOXLINE	562,130	12,800	Monthly
On-line and backfiles	1,225,013		
CHEMLINE	506,271	7,300	Every 2 months
CATLINE	333,800	260	Weekly
CANCERLIT	269,672	3,100	Monthly
HEALTH	186,999	2,500	Monthly

[a] Does not include the 116,620 retrospective records added July to September 1981.

SOURCE: National Library of Medicine.

---

*At this writing, May 1982, NLM offers EPILEPSYLINE, a bibliographic data base produced by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) of NIH from abstracts generated by the private firm, Excerpta Medica. Excerpta Medica provides NINCDS with computer tapes under a contract, at the cost of the tape. The data base is expected to be available through Bibliographic Retrieval Services (BRS) in the very near future, and thereafter will no longer be accessible through NLM. Use of EPILEPSYLINE is funded by NINCDS, and is rather low (fewer than 40 hours per month on average). Both NINCDS and NLM believe the data base will find a larger audience through BRS (159).

Table 3.—MEDLARS: Usage of NLM Data Bases During September 1981

Data base	Number of hours
MEDLINE	5,565
MED 77	968
CATLINE	919
TOXLINE	826
CANCERLIT	398
MEDLEARN	389
HEALTH	318
CHEMLINE	285
SERLINE	162
POPLINE	161
TDB	160
SDILINE	149
AVLINE	137
RTECS	94
MeSH VOCABULARY	86
TOXBACK 74	70
NAME AUTHORITY	42
HISTLINE	38
EPILEPSY	34
BIOETHICSLINE	31
CANCERPROJ	31
CLINPROT	17
STORESEARCH	2
Total	10,882

SOURCE: National Library of Medicine.

many other health-related disciplines. The intent is to provide references to the most useful biomedical literature from the most significant biomedical journals. Author abstracts are included for about 47 percent of the articles cited. Literature selection, thesaurus maintenance, and indexing are three MEDLINE-related activities that require subject matter knowledge. MEDLINE's ability to provide relevant bibliographic references depends on the performance of these activities.

Literature Selection.—Literature selection is used as a quality filter for the indexed biomedical literature data base. About 95 percent of the citations in MEDLINE are those selected for the printed *Index Medicus;* the remaining 5 percent include citations from the *Index to Dental Literature,* the *International Nursing Index,* and a limited number of nonserial publications. Although NLM chooses which journals are to be indexed for *Index Medicus,* it has no formal influence on journal selection for the other indexes.

The selection process is highly structured and involves a critical review of the literature by a panel of expert consultants. Journals are selected on the basis of scientific merit and relevance to NLM objectives. The number of serials indexed for *Index Medicus* is low in relation to the Library's collection: only 2,664 of the more than 20,000 serials collected by the Library were indexed in 1981; the number of articles indexed that same year was 273,750. (See app. E for a comprehensive discussion of literature selection.)

Medical Subject Headings (MeSH).—Central to the development of MEDLINE is its controlled vocabulary, MeSH, a list of subject headings consisting of over 14,000 terms used in the indexing process to characterize an article's content. One way to retrieve citations to articles in *Index Medicus* and MEDLINE is by using these terms. Catalogers also use the MeSH terms to catalog books and other documents in NLM's collection. AVLINE, BIOETHICS, CATLINE, HEALTH, MeSH, POPLINE, and SDILINE are also "MeSH"-searchable data bases.

MeSH is arranged alphabetically and categorically. The 15 categories are subdivided, and arranged in a hierarchical manner, to show relationships between broader and narrower terms. The terms are constantly being updated to reflect changes in knowledge and practice. As of 1981, there were 9,000 major terms and 5,000 minor terms in the MeSH terminology. Both major and minor terms are used to describe articles cited in MEDLINE: only major descriptors are used to describe articles in *Index Medicus.* When an indexer assigns a minor descriptor for on-line searching in MEDLINE, the computer assigns the appropriate major descriptor under which the citation will appear in *Index Medicus.* The MeSH vocabulary is also used by catalogers in libraries throughout the world as subject descriptors for books and monographs.

MEDLINE can be searched not only by the controlled MeSH vocabulary, but also by using any word contained in the title of an article, or in the abstract when included. (For a full description of MeSH, see app. E.)

Indexing.—Subject indexing is a disciplined, intellectually demanding activity that requires consistency and accuracy in assigning subject headings to articles. In part, the effectiveness and reliability of a bibliographic retrieval system are built on the reliability of its data bases, which is determined by the indexing quality of its records. Indexers assign MeSH headings to describe an article's contents (indexing) on the basis of their orientation, training and judgment of the article's major and minor points and the headings' congruence with the subjects discussed in the article. They may modify the MeSH headings with one or more subheadings.

An article may be indexed exhaustively (indepth) and assigned about 10 MeSH terms, or not exhaustively (nondepth) and assigned approximately 5 MeSH terms, depending on the length and content of the article. All the articles in most of the journals covered by *Index Medicus* are indexed. In some journals, however, only selected articles are indexed; that is, since some journals carry health-related articles only occasionally, indexers are instructed to scan these journals and index only relevant articles.

Articles are indexed not only by NLM staff, but also by commercial contractors and by centers in foreign countries with which NLM has quid pro quo bilateral agreements. (See app. A for a more complete description of NLM's international activities.) Of the 273,000 articles indexed for *Index Medicus* in fiscal year 1981. NLM staff indexed 70,000 (26 percent), four U.S. commercial contractors indexed 52,000 (19 percent), foreign centers indexed 28,000 (10 percent) and U.S. commercial contractors indexed 123,000 (45 percent) for the foreign centers. Initially, all of the foreign centers did their own indexing. Currently, centers in the United Kingdom, France, Sweden, West Germany, South Africa, and the Pan American Health Organization Library in Brazil all have indexers on their staffs.

NLM exercises a considerable degree of quality control over the indexing process. In addition to requiring high educational standards for indexers, it provides them with a formal training course and continuous contact with able and experienced indexers, NLM revisors, who monitor the proficiency of all indexers. The Library also periodically updates training, and employs computerized validation routines and proofreading at a number of stages during the indexing process. (See app. E for a technical description of indexing.)

Leasing

NLM leases 12 of its data bases through the National Technical Information Service (NTIS) of the Department of Commerce. Payments from these leases are turned over to the U.S. Treasury after NTIS deducts a brokerage fee. As of January 1982, Bibliographic Retrieval Services (BRS) and DIALOG Information Services, Inc. (DIALOG) had signed such agreements with the Library. Until January 1982, the MEDLINE tape was leased to a subscriber at $50,000 for the first year and $30,000 for each subsequent year. Other MEDLARS data tapes were leased at different fee levels. NLM has established the leasing charges and arranged for and negotiated the agreements. NTIS, which acts as an accounting broker, receives payments from the leases and returns these revenues to the U.S. Treasury after taking a 10-percent brokerage fee; in the case of some data files such as CHEMLINE, NTIS pays royalties to the organizations that compile some of or all of the data file.

As of January 1982, the MEDLINE data base tape is leased on a fixed fee plus a use fee basis. There is a use fee of $4.00 per on-line connect hour and $0.01 per off-line printed citation for MEDLINE and MEDLINE back files, with a minimum fee of $20,000. After the $20,000 minimum is met, the use fees of $4.00 per on-line connect hour and $0.01 per off-line printed citation continue in effect. NTIS continues to serve as NLM's broker.

On-Line Services Users and Uses

In fiscal year 1981, NLM conducted over 2 million on-line searches, fully one-third of all such searches performed in the United States (165). In fiscal year 1977, 754 domestic institutions had direct access to MEDLARS. By 1981, this figure had grown to almost 1,550 domestic on-line centers. Users now include the major medical school libraries, over 700 hospital libraries, and libraries in pharmaceutical and other commercial firms (see table 4) and centers in 13 foreign countries. In all, there are more than 1,890 domestic and foreign institutions with direct access to NLM's data bases

Table 4.—MEDLARS: Number of Domestic Institutions On-Line by Region and Type, September 1981

	Type of Institution										
Region	Medical school	Allied health	Other university college	Other library	Hospital/ clinic	Research	Foundation Society Association	Commercial	Information agency	Other	Total
01	9	3	2	1	45	3	0	21	0	2	86
02	14	7	7	3	55	13	4	82	4	2	191
03	7	5	1	1	35	8	3	28	3	2	93
04	15	4	4	5	51	57	10	44	11	16	217
05	7	5	4	0	104	7	3	35	1	5	175
06	17	10	9	2	52	8	2	12	1	2	115
07	14	9	8	2	95	4	9	36	4	0	181
08	10	8	4	1	54	4	3	15	0	3	102
09	14	10	5	0	43	7	4	19	3	3	108
10	2	4	3	3	37	5	2	2	0	2	60
11	12	5	10	2	113	11	5	34	10	7	209
Total	125	70	57	20	684	127	45	328	37	44	1,537
					September 1979[a]						
Total	120	NA	91	NA	400	93	NA	273	NA	34	1,011

NA = not applicable.

[a] During 1980, classifications were expanded.

SOURCE: National Library of Medicine.

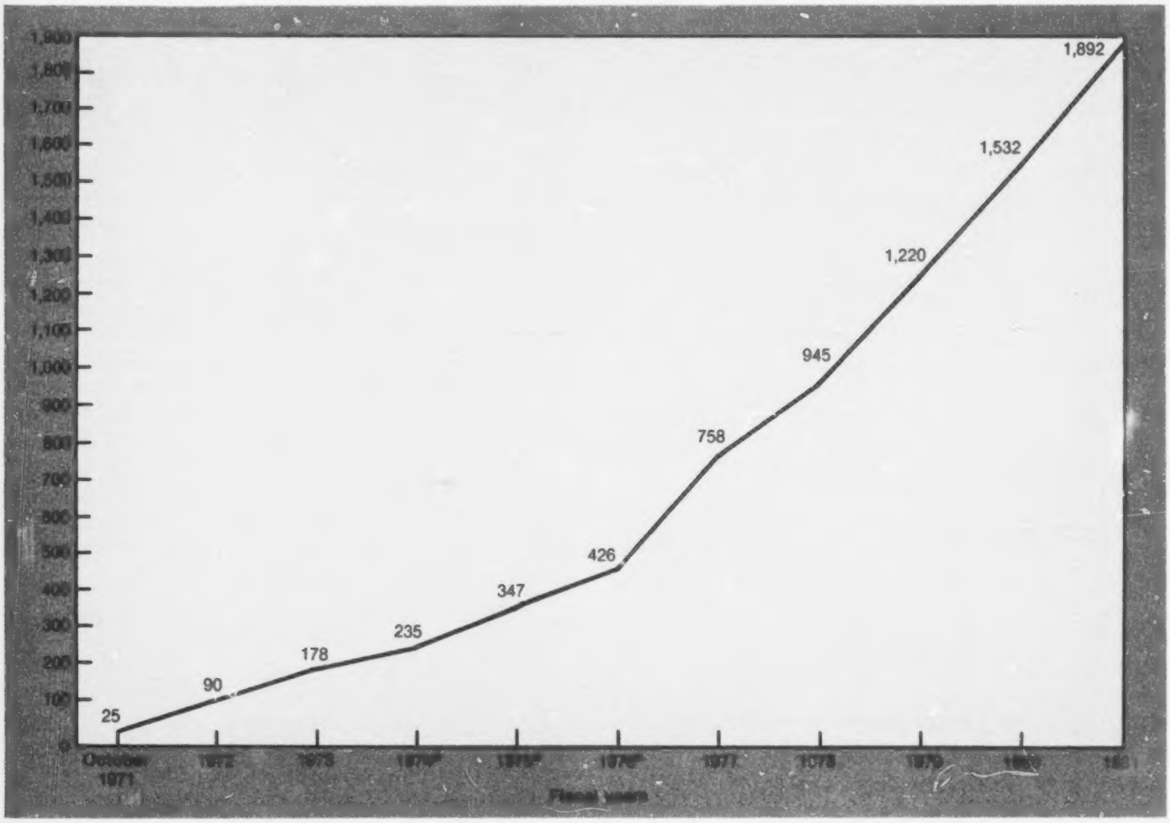
(see fig. 1). Many other institutions, including 3,000 to 4,000 hospital libraries in the United States, provide indirect access to MEDLARS by referring requests to facilities with on-line terminals for searching.

Although NLM does not collect data on the individual users of its services, some information is available on institutions with direct access to MEDLARS. In fiscal year 1981, the Library ran 2.02 million search requests. Of these, 32 percent came from hospital on-line centers, 21 percent from medical schools, and 12 percent from commercial firms (see fig. 2). Hospitals, more than any other type of institution, have direct access to MEDLARS. In 1981, they were the largest users of MEDLARS as measured by both the number of searches requested (32 percent) and the number of computer connect hours (34 percent). Medical schools ranked second, with 21 percent of searches and 17 percent of computer connect hours. Hospitals and medical schools also ranked first and second, respectively, in their utilization of the MEDLINE and HEALTH data bases (see fig. 3).

There are 328 domestic commercial firms with direct access to MEDLARS. Commercial firms are the third highest user group in the utilization of all MEDLARS data bases, as measured by searches performed (11 percent), and the fourth highest when measured by number of connect hours utilized (12 percent) (see fig. 2). Ten percent of the MEDLINE connect hours are used by commercial firms (see fig. 3). These firms tend to use CHEMLINE and TOXLINE more often than MEDLINE and utilize an estimated 30 percent of the total connect hours to TOXLINE.

There are no recent nationwide data on the ultimate users (i.e., end users) of MEDLARS data bases. Most studies of MEDLARS users have design limitations, are dated, or have been conducted on an institutional, local, or regional base. They indicate that a variety of health professionals, including students, request searches for a variety of reasons (see app. C). For example, a survey conducted by NLM in 1975 found that 41 percent of MEDLARS end users were physicians, 19 percent were nonphysician scientists, 10 percent were librarians, 10 percent were students, and 20 percent were reported as "other" (85). Preliminary results from a survey of hospital on-line centers, mainly located in New England, indicate that slightly more than one-half of search requests are from nonphysician health providers, including nurses, ancillary service providers, and administrators. Searches are requested to aid in providing patient care, preparing presentations and journal articles, and planning new services and purchasing equipment (50).

Figure 1.—MEDLARS: Number of Domestic and Foreign Institutions On-Line, 1971-81



aFigures do not include users with TOXLINE file access only; therefore figures are lower than actual.

SOURCE: National Library of Medicine.

Searching and Retrieval

MEDLARS II allows for direct communication with the computer in on-line, interactive fashion. That is, the user can carry on a "dialog" with the computer, refining the search by typing in successive questions until the needed references are identified. An on-line search usually takes 10 to 15 minutes. (See app. E for a description of searching methodology.)

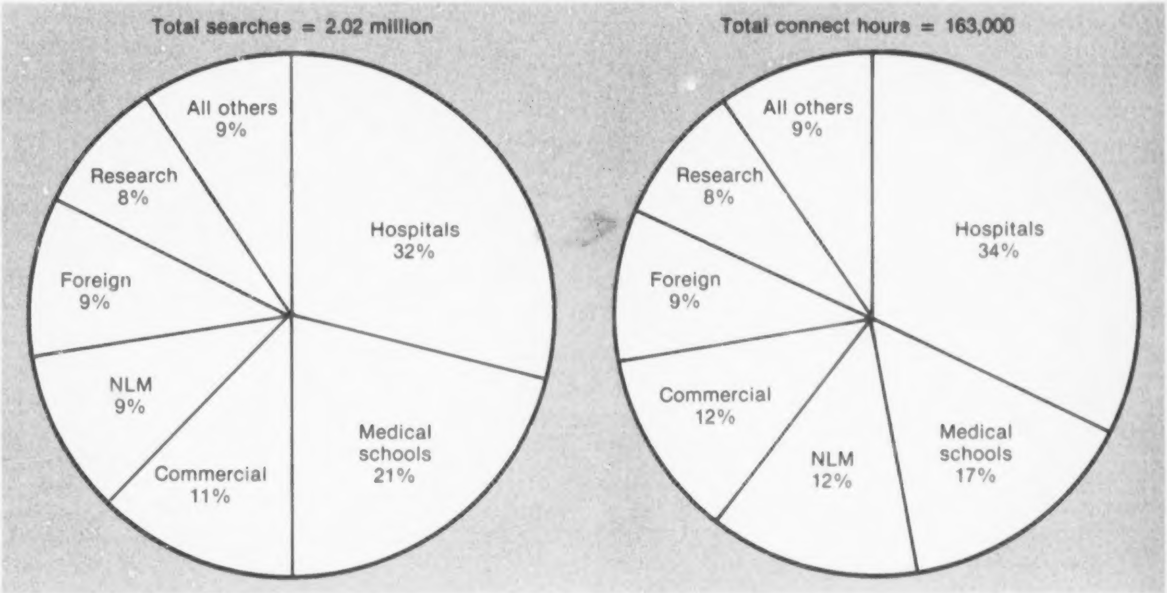
The references can be either printed at the time the search is entered (on-line) or printed during the offpeak hours (off-line) and mailed to the requester from NLM or SUNY the next morning. (Searches can also be formulated on-line and then stored in the computer for reference and later use.) The number of on-line searches more than doubled from 1977 to 1981 (see fig. 4).

For the most part, trained search analysts, termed intermediary users, perform searches. Although an untrained individual can perform searches, the current system, like most others currently in use, is not designed for it. NLM is investigating methods to make the system more cordial to users without library or information systems training (64).

It is generally agreed that an ideal search requires that a trained searcher have a reference interview with the person requesting information

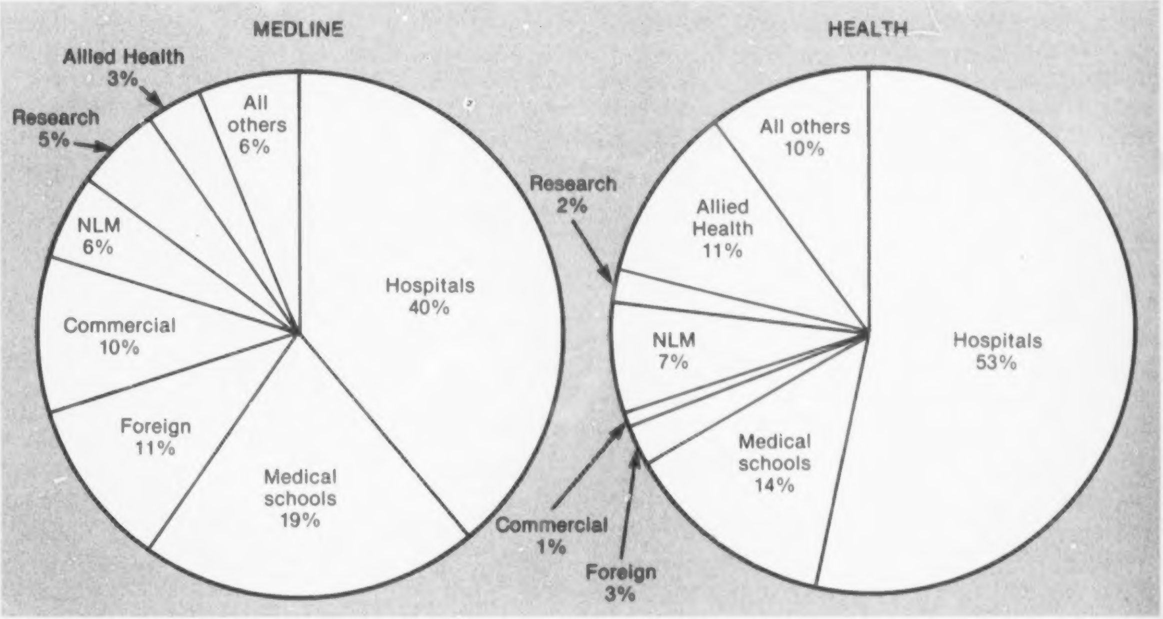
BEST AVAILABLE COPY

Figure 2.—MEDLARS: Percent of Searches and Connect Hours by User Categories, Fiscal Year 1981



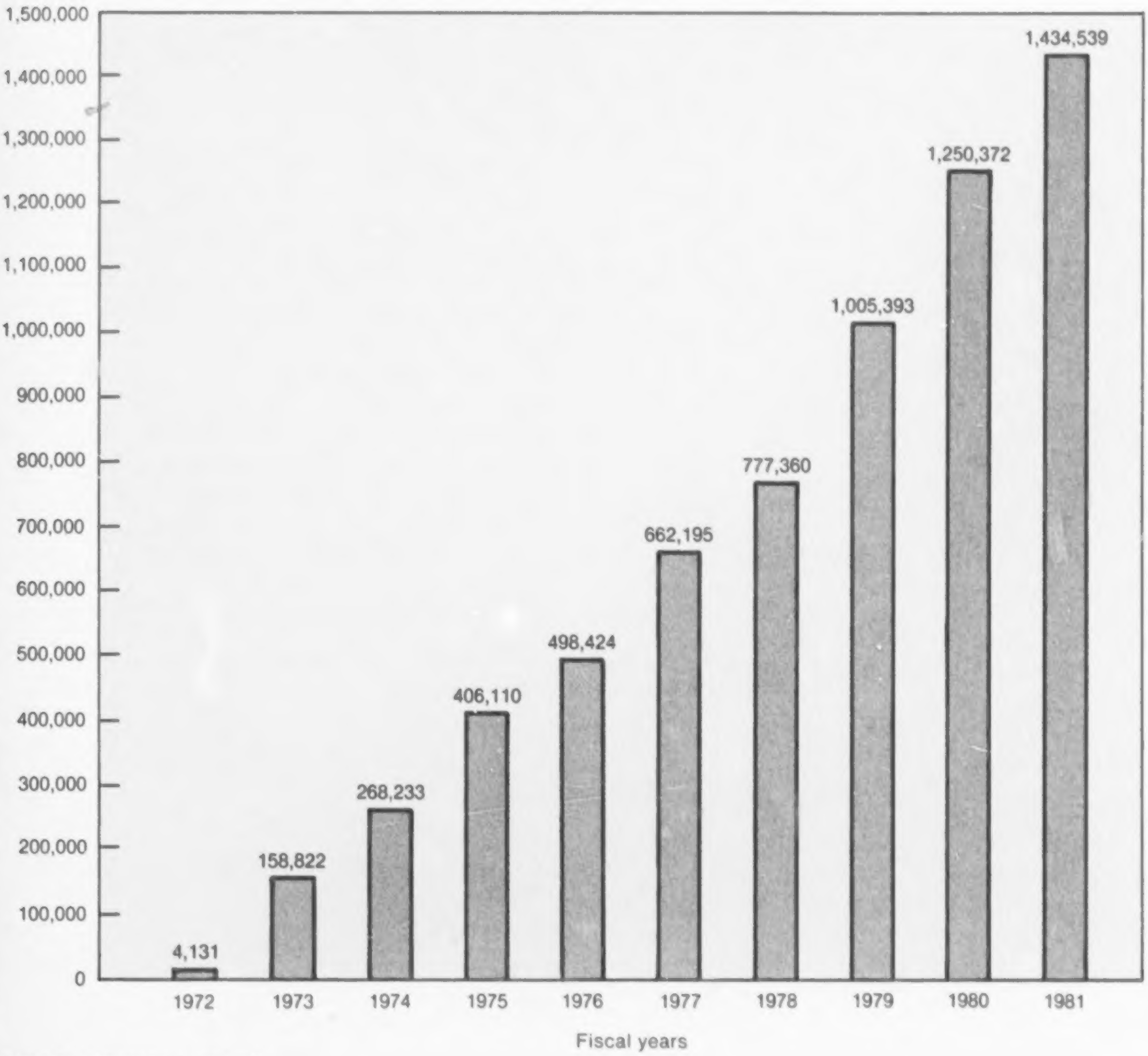
SOURCE: National Library of Medicine.

Figure 3.—MEDLARS: Use of MEDLINE and HEALTH On-Line Data Bases by Connect Hours by User Categories, Fiscal Year 1981



SOURCE: National Library of Medicine.

Figure 4.—MEDLARS: Number of On-Line Searches Performed, 1972-81



SOURCE: National Library of Medicine.

to determine the purpose of the request, the sources the requester has consulted, the facets of the subject the requester wants emphasized or eliminated, new terms in the field the requester can supply, and the character and volume of the retrieval the requester expects. Requesters are often familiar with such information from their own experience and from previous consultation with written sources. It appears that not all search requests are conducted in this manner, resulting in a wide variation in the quality of searching results.

Aggregate data on how search requests are made are not available, but there appears to be variation from library to library, mainly because of differences in populations served. For example, the New York Academy of Medicine Library

BEST AVAILABLE COPY

estimates that more than 80 percent of its searches are requested by telephone or mail. Most of its other inquiries are by telephone. The reference room of NLM estimates that about 50 percent of its research requests are made in person, while the Biomedical Library at the University of California at Los Angeles, an academic health science center library, reports that over 80 percent of its requests are made in person.

User Training

NLM has developed an on-line training course in MEDLARS for information specialists who act as intermediary searchers for end users. The first courses, held in the mid-1960's, initially trained analysts in batch searching techniques during a rigorous 8-month session; this was later decreased to 2 months. With the introduction of interactive on-line searching in October 1971, the need for lengthy training decreased, and the training program was shortened to 3 weeks. The current search training program lasts only 1 week. Before attending the formal 1-week course, a searcher must complete MEDLEARN, the Library's computer-assisted instruction program.

The 1-week training course, normally held either at NLM or at the Regional Medical Library in Los Angeles or Omaha, is comprehensive and includes didactic instruction in system mechanics, Boolean logic, search formulation, controlled vocabulary searching, free text searching, the scope and content of the MEDLARS data bases, special system capabilities, as well as hands-on experience at the terminal. The training course is set up in a modular fashion, in order to assist searchers who need more complex and more comprehensive training. Additional training is available after the initial class, and many searchers return for a 1-week advanced training course. NLM also teaches abridged update/review courses.

During fiscal year 1981, 942 search analysts were trained in 43 initial and advanced training classes at NLM, UCLA, the University of Nebraska, and in the field. This is more than triple the number (254 analysts) training in 1977.

As noted previously, more than 1,890 domestic and foreign institutions now have direct access to MEDLARS via NLM. New institutions have been admitted to the MEDLARS network, as computer capacity permits, on a priority basis. Priority standing is given to direct patient care facilities, health professional education institutions, organizations primarily engaged in health protection activities, and Federal and State health agencies. As a condition of obtaining access to MEDLARS, an organization is required to send at least one person to attend the formal 1-week initial training program. Salaries and other expenses during training must be paid by the sponsoring organization.

In the past few years, more hospitals/clinics and commercial institutions have obtained direct access to MEDLARS. As a result, the character of the student population has become more heterogeneous. The original analysts were experienced librarians from large medical schools, hospitals, research institutions, and Federal agencies, resulting in a fairly homogeneous class. Today, the varying levels of expertise represented in the classes have stimulated NLM to experiment with new training methods and curricula.

Notwithstanding the mandatory searcher training program, there is some opinion that searcher variability may well be the weakest link in the entire system. The MEDLARS system may be too complex to be understood in a 1-week course by one who has little background in information systems or medicine and biology. Since there are neither entry nor exit standards for trainees, there is no control over the level of expertise of those who attend the sessions. The trainees' difficulty is mainly in the complexity of the subject, the data base content, and the thesaurus. NLM training provides searchers with highly technical skills, yet the knowledge base grows ever more complex, and thus, so must searchers' skills. In addition, once a terminal is acquired by an institution or organization, anyone can use it.

The interactive process between the searcher and requester is critical to identifying relevant citations in the data base. Inexperienced and poorly trained searchers cannot respond adequately to such a need. Searchers must meet the needs of practicing physicians who tend to use MEDLARS to obtain information related to clinical problems.

These physicians, often not accomplished researchers, are also the ones most in need of guidance to ask the right questions. The critical point is whether the searcher really understands what the user wants. Such issues become more important as NLM extends MEDLARS throughout the hospital community. At many places in the Nation, including the University of Virginia, a concerted effort has been made to upgrade the skills of librarians in community hospitals remote from major medical centers and large resource libraries, so that they may adequately conduct search interviews with physicians and forward information requests to resource libraries.

There are some observers who suggest that problems with retrieving information from the system may not be particularly attributable to the searcher, and they point to an evaluation study of MEDLARS' on-line process conducted in 1978-79 (156). That study concluded that the type of training (formal or informal) had no statistically significant effect on search performance, and suggested that in some areas improving search effectiveness may be more a matter of system redesign than of training or retraining. The study also found the predominant reason searchers elected to use one system over another was that the system they elected offered access to more of the data bases that they needed. In addition, searchers tended to prefer the system they had learned first. Searchers were also influenced by the range of a system's capabilities.

User Services

In addition to providing search training, NLM supports the users of MEDLARS in numerous ways. In order to facilitate interaction between NLM and user institutions, a professional library staff member is assigned to a user service desk at NLM to receive telephone calls and respond to problems which users are having with the system. In addition, information about the system is transmitted to enrolled users by newsletters and other forms of communication. When a new MeSH is published, and when new on-line data bases are added to MEDLARS, the regional medical libraries update the training of the searchers in their regions. Searchers also have access to the *NLM Technical Bulletin* and *Users Manual* and may contact a more experienced searcher for assistance.

On-Line Access Charges

NLM began charging users in the United States for access to its data bases in April 1973. The prices have increased a number of times over the years. As of January 1, 1982, computer connect time for all data bases, except CHEMLINE, TOXLINE, TOXBACK 65, and TOXBACK 74, is $22.00 per hour of prime time (10:00 a.m. to 5:00 p.m. EST) and $15.00 per hour of nonprime time. The price for printing a page of citations off-line is $0.15. The price of CHEMLINE is $101.00 for each hour of prime time and $94.00 for each hour of nonprime time, with a charge of $0.45 for each page printed off-line. TOXLINE and its backfiles are priced at $52.00 per hour of prime computer time and $45.00 per hour of nonprime time, with a charge of $0.35 for each page printed off-line.

CHEMLINE, TOXLINE, TOXBACK 65, and TOXBACK 74 are priced to include the royalty charges paid to the organizations that provide information for them (e.g., Chemical Abstract Service, BIOSIS, and the American Society of Hospital Pharmacists). Foreign centers establish their own rates.

NTIS serves as a collection agent for NLM. In 1980, roughly $2 million in charges were paid by domestic user institutions in on-line access charges. NTIS paid 70 percent of this sum to contractors supplying NLM with telecommunication systems and additional computer capacity; 20 percent of the charges went for billing, collection, mailing, and postage; 10 percent was returned to the Treasury.

THE FUTURE OF MEDLARS

Current Standing

The developmental efforts of NLM that led to MEDLARS II promoted the development of on-line systems by both the public and private sectors. The private sector took the opportunity and studied the NLM system: "Tests and analyses of the system and its use had a significant impact on information science researchers and helped guide their thinking along new lines" (165). NLM had acted in the traditional role of Government by demonstrating the application of novel and potential technologies that require large and long-term investments not typically assumed by private industry.

Many of the on-line information services that were formed in the wake of NLM's developmental efforts have since revised and expanded their search capabilities more rapidly than NLM, and MEDLARS II is no longer a state-of-the-art system.

In May 1980, the Director of NLM established a MEDLARS III Task Force to plan for NLM's future automation needs. The Library's computer experts are troubled that the functional life span of the MEDLARS II software is only 3 or 4 more years. All software eventually becomes "fragile" because of the continuous modifications required. The search services MEDLARS II provides are not as sophisticated as those available on other on-line systems.

More important, many of NLM's internal operations are now carried out through a combination of manual techniques and separate automated systems, and the Library is finding it difficult to continue providing quality services without enlarging its staff. Since such an increase is not likely in the immediate future, NLM hopes that more and improved automation of internal functions will prove compensatory. New techniques for computerizing internal library functions are available and in use in other libraries.

The MEDLARS III Task Force completed the conceptual framework and the functional analysis of MEDLARS III in September 1980. The systems development plan was completed in September 1981. Currently, a systems analysis is under way, with a request for a proposal to design and implement MEDLARS III expected in 1982. The new system is expected to be completed incrementally between 1983 and 1985. The system analysis and systems development plan are being prepared by NLM staff supplemented by specialists from the private sector. Current plans for the design and implementation of MEDLARS III are that both tasks be carried out by a commercial firm under contract.

MEDLARS III is a line item in the NIH/NLM budget and is expected to cost $6 million. The functional analysis component cost $1 million; the design, scheduled for 1982, is expected to cost $2 million; the design and implementation, scheduled for the year after, will be funded at $2 million; and the installation, scheduled for the last year, is funded at $1 million.

The MEDLARS III Task Force was advised by experts and consultants from both the health sciences library community and the general information systems community in both its analysis of the Library's needs and in its formulation of functions to be incorporated into MEDLARS III. The task force expects that MEDLARS III "when implemented [will] represent a significant transition from a partially automated, loosely integrated set of functions [characteristic of MEDLARS II] to a highly automated, tightly integrated system that is responsive to the needs of many" of its potential users (104). This statement reflects the pace at which technological advancements are moving in this field.

Generally, MEDLARS III is expected to improve NLM's internal and external capabilities. It will automate and improve: 1) the acquisition of biomedical literature; 2) the creation, maintenance, and distribution of bibliographic records; 3) the retrieval of bibliographic information; 4) inventory control; and 5) the delivery of primary documents (i.e., interlibrary loans). Some of these processes are now automated, whereas others are essentially manual. MEDLARS III will not perform functions that are purely commercial and available from private vendors, such as putting the user's own private files on-line for searching.

Improvements in MEDLARS' retrieval capabilities will result from the construction of an integrated data base that permits a single search access to the bibliographic data bases currently in the system. There will be other features in MEDLARS III to assist and simplify a search analyst's efforts. It will adhere to standards for forming search strategies now being established by the Council of Library Resources' Committee on Man-Machine Interface.

MEDLARS III is designed to build on MEDLARS II, permitting the incorporation of new technology and allowing for incremental implementation. For example, NLM has a contract to explore the initial phase of on-line indexing. As soon as this indexing is achieved, it can be incorporated into the old system until MEDLARS III is ready. Much of the new system will not be at the leading edge of computer technology, but will update the current capabilities of MEDLARS II and incorporate newly established or developing technologies. Electronic mass storage and distribution of textbook material is being studied at the Lister Hill National Center for Biomedical Communications, and an experimental retrieval language, CITE (Current Information Transfer in English), which should make the system more accessible to an untrained searcher, is under development at NLM.

The plans do allow some latitude for examining new and undeveloped innovations. One innovation that could be considered for MEDLARS III is a natural-language approach to computers (i.e., involving the use of standard English) that was developed at Beth Israel Hospital in Boston, under a research grant from NLM. This approach has been very successful in a limited system, and NLM is considering funding a feasibility study for enlarging the system to include all the hospitals in the Boston area, hoping that the results may be applicable to MEDLARS III. The purpose of using natural language is to make the system easier for untrained people to use. A comparison of features between MEDLARS II and MEDLARS III is presented in table 5.

At the same time that MEDLARS III is evolving, NLM administrators are considering the future of health information transfer. The technology and the times may warrant the introduction of more distributed kinds of arrangements for the provision of data base information (see app. I), and NLM has undertaken a limited experiment with the Toxicology Data Bank to test this hypothesis (33).

Table 5.—Summary Comparison of the Functional Capabilities of MEDLARS II and MEDLARS III

MEDLARS II

Retrieval: Highly automated, oriented towards the technically trained search analyst.

Technical Processing—Level of Automation: Many critical functions such as serials processing, indexing, catalog maintenance, circulation, and an interlibrary loan are essentially under manual control.

Record and File Creation and Maintenance: Currently automated functions have deficiencies which include duplicate redundant record creation, maintenance, and searching; inclusion of nonauthoritative bibliographic data; and incompatibility between NLM files and with other national library data files such as the Library of Congress, Ohio College Library Center.

Card Catalog: Manually controlled; difficult to maintain up-to-date.

Network Access and Interface: On-line access to NLM records is available only for the retrieval. Only limited access to capabilities such as record creation and maintenance. No capabilities to provide for local records creation.

Regional Medical Library Document Delivery System: Manual system for processing requests, maintenance activity and administrative controls for almost 1,000,000 requests per year.

MEDLARS III

Retrieval: Enhanced user aids with natural English-like language queries will benefit not only the trained searcher, but will enable health practitioner users to access the system directly and effectively for many queries.

Technical Processing—Level of Automation: These functions will be more completely automated and integrated.

Record and File Creation and Maintenance: Will be based on the concept of a master machine-readable record for each bibliographic entity and processing unit. Record structure will be based on the Library of Congress MARC record. All bibliographic elements will comply with established standards.

Card Catalog: Fully computerized and available on-line through MEDLINE network to users all over the United States.

Network Access and Interface: Users will have access to national locator information and the retrieval system files. The system will be capable of interfacing on-line with other automated systems to provide data for support of local processing, or to permit distributed processing of NLM records in local systems.

Regional Medical Library Document Delivery System: Fully automated system for generating and routing requests. Fully computerized management control of traffic including fiscal and program control.

SOURCE: National Library of Medicine.

Chapter 3

MEDLARS: System-Related Issues

Chapter 3

MEDLARS: System-Related Issues

INTRODUCTION

The only full-scale effort to review the effectiveness of the MEDLARS service was conducted in 1966-67, before the system was available on-line. At that time, very little was found to be critically wrong with the system (75). The study was conducted on an earlier version of MEDLARS, in essence examining a system vastly different from that in place today. (See app. C for a review of evaluative studies of MEDLARS.) A comparable examination of MEDLARS at this time would require a much greater effort and may be unwarranted.

From a review of the literature, interviews with trained users of the system, consultation with experts, and statistics on system utilization (see ch. 2), *OTA finds that MEDLARS, in general, is effective in disseminating health information, using the traditional criteria of recall and precision.** The system does not retrieve information as efficiently as is possible with available technologies, but its retrieval and other capabilities will be enhanced when its successor, MEDLARS III, is developed.

OTA has reported on issues concerning MEDLARS' effectiveness in disseminating bibliographic health information in part in the staff paper "The National Library of Medicine" (116) and in part in the OTA report *Strategies for Medical Technology Assessment* (117). This chapter considers three system issues: 1) the subject content of the literature cited in the MEDLARS data bases, in particular MEDLINE; 2) the coverage of nonserial literature in the MEDLARS data bases, in particular MEDLINE; and 3) the evaluation of the methodological design of articles in literature cited in the MEDLARS data bases.

**Recall* refers to the ability of an information system to retrieve "relevant" documents; i.e., documents of value in relation to an information need that prompted the request for an on-line search. *Precision* refers to the system's ability to hold back "nonrelevant" documents (75). These two measurements are considered to provide (by implication) an estimate of a system's ability to satisfy the information needs of its users.

SUBJECT COVERAGE

The National Library of Medicine (NLM or the Library), like other organizations responsible for collecting and organizing biomedical literature and for providing information services, is faced with the difficult issue of delineating the field of biomedicine. Over and above the tremendous growth in the quantity of published biomedical literature, the boundaries of traditional medicine as a field of practice and research are continuously expanding to encompass new disciplines. Many disciplines are fragmented and becoming more and more specialized, and new interdisciplinary fields are created in the pursuit of scientific knowledge.

The fields crucial to the health of the Nation's people have become so numerous and diverse that discriminating between those central and those peripheral to health depends on one's point of view. Conventionalists maintain that the basic medical sciences and clinical medicine disciplines are at the core of health and should remain the primary focus of the Library's efforts. Others—who consider that social, psychological, and environmental factors influence health, or who are interested in the application of research findings or in the organization of health services and its effect on health—disagree. They propose that areas such as behavioral medicine, technology assessment, primary care, family medicine, and health services research are equally important, but are receiving insufficient or delayed attention by the Library for its collections and in its products and services such as MEDLARS.

NLM has tried to be responsive to all users and has expanded its subject coverage beyond the fields historically associated with biomedicine. In 1879, Billings classified the citations in *Index Medicus* into 14 categories, one of which, jurisprudence, may be considered outside the medical model. In 1973, there were 81 fields or disciplines used to classify *Index Medicus* citations, with many new entrants from disciplines outside the basic sciences and clinical specialties of medicine (31).

One example of NLM's efforts to expand subject coverage is in its coverage of the health services literature. In 1966, NLM, at the request of the American Public Health Association (APHA), began to broaden its selection of journals and to expand its set of subject headings to index health services literature. APHA established a committee to advise NLM "in its efforts to improve the analysis, storage, and retrieval of literature pertinent to . . . medical care organization and financing" (105).

This effort, while important, was apparently insufficient. A 1976 report by the House Committee on Interstate and Foreign Commerce found that NLM had not adequately served the information needs of those concerned with health care delivery and health services research, and that attempts to retrieve information in the fields related to health care delivery had not been entirely satisfactory because of deficiencies in the vocabulary used for indexing and cataloging (149). The report did, however, commend the Library as "a world leader in library services for medical sciences and biomedical research" (149).

NLM renewed its efforts to become a primary source of information to professionals in health services research and health care delivery by convening additional advisory groups, entering cooperative arrangements with the American Hospital Association, and continuing to expand its vocabulary and the number of journals indexed for MEDLARS concerning health care delivery and health services research.

By 1977, 58 serial titles, recommended by organizations and individuals in health care management, economics, law, and manpower, had been added to MEDLINE. The Library was also adding indexing terms to its medical subject headings (MeSH) vocabulary: 150 by 1977, and an additional 50 to 100 in 1978. A 1977 House Interstate and Foreign Commerce Committee report congratulated NLM for making "substantial progress" in handling health care literature (148). By November 1978, the Library had established a distinct data base, HEALTH, that contained references to literature on health planning, organization, financing, management, manpower, and related subjects. The American Hospital Association assists with the updating of the data base.

The Library also began a collaborative effort about 5 years ago with the National Health Planning Information Center (NHPIC). In 1979, NHPIC started working with the Library to develop a common terminology, so that NHPIC's data base could be included in NLM's HEALTH data base. Serials from NHPIC's data base have been included in HEALTH since 1979. It is expected that by the end of 1982, the nonserial literature in NHPIC's data base back to 1975 will be indexed according to MeSH and will be available on HEALTH. While terms are continually added to MeSH, health services researchers and planners find some information retrieval problems that mainly stem from an absence of clear definitions and distinctions among indexing terms and from the inconsistent application of terms by NLM indexers in preparing HEALTH (24). Both factors are normal indicators of a relatively new and changing field.

Coverage of health care literature will probably remain a problem for NLM, because relevant articles and literature appear in so many diverse publications and reports. Selectively indexing more journals alone from law, management, and public affairs would be exceedingly difficult and expensive. Furthermore, NLM's experience with selectively indexing health planning administration journals since 1976 indicates a skewed distribution for the selected articles: a very few serial titles contain the vast majority of the relevant material. Improved coverage would require reviewing a larger number of journals containing few relevant articles per issue (24).

NLM continues its attempts to improve the subject coverage of many other new and emerging fields. The Library is now concentrating on improving the literature collection and the MeSH vocabulary for the psychological and sociological aspects of medicine. But, according to certain observers and users, some other fields essential to health such as primary care are being overlooked by NLM. Because similar objections are being raised about NLM's coverage of the nonserial literature, these two issues will be considered in the discussion section of this chapter.

LITERATURE COVERAGE

An issue closely related to the issue of changes in the character of the content of biomedical literature is that of changes in the type and form of the literature. The format for presenting information often varies among disciplines. For example, in addition to books and serial publications, the health care literature, perhaps more than the literature in other biomedical disciplines, often includes unpublished technical reports, project descriptions, speeches, and presentations that are referred to as "fugitive literature" or "grey literature." All biomedical fields have seen a rapid growth in such literature in recent years. Strictly speaking, "fugitive" or "grey" literature is not subject to publishing and reviewing channels because of length, degree of detail, specialized language, or restricted interest, according to Public Health Service criteria (76). The term has also been used to refer to literature so widely dispersed in so many sources that it is difficult to find, or to literature that appears in channels not normally expected (e.g., an article on palliative care in *Architectural Forum)* (76).

Indexing services have traditionally covered the "fugitive" or "grey" literature only selectively, with NLM being no exception. MEDLINE, the largest and most extensively used data base in MEDLARS, concentrates on journal literature: 5 percent of its citations were from other literature, primarily proceedings of conferences and symposia. As of January 1982, MEDLINE included references only to journal literature. CATLINE, which includes citations for all post-1801 printed books and serials, contains citations for many published proceedings and theses as well as monographs. As noted in the previous section, by the end of 1982, NLM's HEALTH data base will include the nonserial literature in NHPIC's data base back to 1975, indexed according to MeSH.

As this country's biggest generator of information, the Federal Government through its various agencies and contractors uses technical reports and other unpublished documents as a means of communicating research and development progress. In many cases, such literature is the only communication link from the time a research project is initiated until its results have been formally published in a book or journal. The results of many research projects are of current interest only, or serve exclusively as a tool for furthering a more comprehensive research effort. In the latter case, results may never be published, making the "fugitive literature" the only source of information (152). Some believe that if information is worthwhile, it will in time be published in a journal, particularly if the information is about biomedical research. Others disagree with this view.

The intent of Congress over the years, however, has been to see that the professional and taxpaying publics receive information benefits from the technical missions that it authorizes (3). As a result, the clearinghouse function—to evaluate, package, and distribute unpublished biomedical information widely but selectively—has been used increasingly as a response to the information explosion phenomenon over the last decade. Clearinghouses typically identify, select, acquire, process, and store documents and other materials while providing "locator tools," such as indexes, to this collection. In the United States, there are 41 clearinghouses with a health focus, the majority funded by the Department of Health and Human Services (DHHS) (12).

The Government Printing Office releases about 10,000 Government reports annually. The ultimate processor and repository of federally sponsored scientific research, development, and technical reports is the National Technical Information Service (NTIS) of the Department of Commerce. It has its own bibliographic data base (NTIS) that is available from commercial information services. About 10 percent of the NTIS document collection, which exceeds 1.2 million titles, comes from DHHS, indicating a substantial flow in health-related information to NTIS. Furthermore, NTIS has created working relationships for the computerized preprocessing of documents with at least three entities within DHHS: the National Cancer Institute, Project Share Clearinghouse, and NHPIC.

METHODOLOGICAL MERIT

The expansion in biomedical publication and the diversification of biomedicine in subject and format has been accomplished, as noted earlier, by a technological revolution. With massive increases in the storage capabilities of computers and improvements in communications systems, the volumes of information and data that can be accessed are overwhelming. A problem that is becoming more and more significant is that of "information overload" and the need for readers of the literature to separate the wheat from the chaff. The concept of quality filtering, which was first introduced by Etzioni in 1964 (47), has received attention from a number of investigators (119,136, 157) and at international conferences (30,124).

The methodological design and the statistical analysis used in many articles, even in prestigious journals, may be questionable. For example, of 67 clinical trials reported in 1979 and 1980 in the *British Medical Journal, Journal of the American Medical Association, Lancet,* and *The New England Journal of Medicine,* only 12 percent reported on the statistical power of the investigation (39).

It appears that journal editors who have acted as information gatekeepers of the scientific community are unable to continue filling this role— in part because of the growing complexity of scientific literature, in part because of the climbing standards of statistical adequacy. Journal editors have been considered winnowers of the scientific literature, "who, with the aid of peer review, sift the finest grains to assure that studies published in the scientific literature are well designed and scientifically and ethically sound and that the findings are valid and thoroughly explicated and that the work constitutes a true contribution to scientific knowledge" (44).

The referee system used in the process of selecting articles by journals is also open to question. It has been shown, for example, that the concurrence between two referees of each of some 500 papers submitted to *The New England Journal of Medicine* was only slightly better than chance (67). In addition, the cost of the referee process in the review of journal articles is high because of the need for input from subject experts (123).

It has been suggested that NLM, in preparing its data bases for MEDLARS, assist with the gatekeeper role by describing articles as to adequacy and appropriateness of the statistical and epidemiological aspects of the articles' experimental design and analysis. NLM currently performs some quality control in the selection of materials for the Library's collection and in the process of selecting literature to be indexed for *Index Medicus* and MEDLINE. However, the selection is based on the scientific merit of the journal as a whole and not on the quantitative accuracy of specific articles, although the merit of individual articles contributes to the choice of the literature.

Rigorous evaluation by NLM of the quantitative methodology used in specific articles would be extremely costly: it would require an increase in the Library's funding as well as an effort to locate and hire of personnel with the requisite expertise, neither of which seems realistic at a time of fiscal retrenchment. Further, it would delay the entry of references to published material into the bibliography (*Index Medicus* or MEDLINE). According to NLM, quality control of journal articles by the Library also "would unquestionably involve substantial debate about some articles where statistical issues are themselves unsettled among experts. Finally, the filtering of published articles puts the Government in the position of a scientific censor, with all the unpleasant implications of Big Brotherism and excessive 'regulation' " (98).

The most advantageous points in the library process and elsewhere for filtering the literature are unknown and warrant serious investigation. NLM could well serve as the catalyst for research in this area, which is certainly germane to its mission. For the time being, NLM could provide minimal guidance to users by providing simple indications about articles (e.g., whether an article has data arrays in such formats as tables and graphs) without making any definitive value judgment as to their merit, or by refining the use of its present methodological subject headings.

DISCUSSION

Expanding the limits of MEDLARS with respect to the subject scope and type of literature covered in its data bases has been discussed in the past, but the funding to accomplish this goal was not forthcoming (65). NLM must operate within financial and personnel constraints. If it remains necessary to contain the perimeters of MEDLARS, a reordering of selection priorities might be conceivable in light of the changing boundaries of the biomedical field and other fields that benefit the public health. But the lack of general agreement as to the relative health contributions of each field is a serious deterrent.

The issue of defining the fields of relevance to health is not unique to NLM; it has been debated in many forums over the years. The issue remains unresolved because virtually every aspect of human culture has some relevance to health. The burden of defining the fields essential to health cannot be assumed by NLM, but the issue is of importance to the Library because its mandate is open to interpretation. As noted in the Library's originating legislation, the purpose of NLM is to "assist the advancement of medical and *related sciences* and to aid the dissemination and exchange of *scientific and other information important to the progress of medicine and the public health*" (Public Law 84-941).

Thus, it would be helpful to the users of MEDLARS if NLM's Board of Regents, within the limits of the statutory language, were to define precisely the scope of subjects and type of literature to be included in the Library's collection and products. Although information on NLM's policy concerning subject coverage and literature coverage is available, the policy is not known by all MEDLARS users. Those interested in fields outside the basic sciences and clinical medicine are sometimes uncertain about the dimensions of the MEDLARS data bases, and some users are not always sure about the definitiveness of their search results. A more interactive mode of communication among the library, librarians and other information specialists, and the ultimate user would enhance the understanding of the situation. Another party at interest is the private sector of the information industry, where perceived inadequate communication about the limits of NLM's data bases and NLM's plans for their modification affects operational and investment decisions (166).

MEDLARS cannot technically or financially cover all aspects of all health-related fields, or comprehensively cover all literature, published or "fugitive," in all health-related fields. Nonetheless, members of various health fields, particularly in new and emerging areas, have expressed a need for better bibliographic access to information of interest. NLM has assisted some professional organizations in the development of new bibliographic products. For example, the Library currently cooperates in the production of the *Family Medicine Literature Index* (*FAMLI*), an index to the international literature in family medicine. NLM decided against increasing the coverage of family medicine magazines to be indexed for *Index Medicus* and for the MEDLARS data bases or creating a special list of the journals. Instead, NLM prepares a recurring bibliography on family medicine from the data base which produces *Index Medicus.* The bibliography is incorporated into *FAMLI* and supplemented by the publishers with references to non-*Index Medicus* journals, which are indexed by the *FAMLI* staff, using MeSH and additional family medicine subject headings (52).

The *FAMLI* model is one of many that NLM could use for assisting health-related professional organizations improve bibliographic access to the literature in their specific fields. A Library policy that would permit the leasing of part of the MEDLINE data base and would permit the reproduction of the data base tape would be a way of developing new data bases of specialized interest. The publishers could then supplement MEDLARS data bases with information they deemed necessary, perhaps including "fugitive literature" citations or additional information subject headings in their subject fields.

The problem of constructing MEDLARS data bases in response to user needs is compounded by the absence of sufficient reliable data with which to construct a user profile. NLM is moving to obtain more data on its institutional user community in its new pricing policy, which requires the payment of a use fee every time one of its data bases is accessed through a commercial information service or a foreign center that leases MEDLARS tapes (see ch. 2). At present, it collects data only on institutions that access its data bases directly through NLM and State University of New York (Albany) computers. NLM needs information about the individual user. For example, data on the end user's profession would be helpful in developing data bases more truly reflective of user needs. Even so, the nature and needs of the user community are changeable, and potential users are difficult to identify.

If MEDLARS data bases were to include an assessment of the methods and statistics used in the articles it cited, the system would be extended beyond its current capabilities. Although the issue of assessment reflects users' needs, it is not specific to MEDLARS data bases only, but is important to all health-related data bases. As noted previously, it may be appropriate for NLM to use its research capabilities to explore this problem.

NLM has already started research on one method of filtering information in the construction of its Hepatitis Knowledge Base. The contents of the data base are not bibliographic references, but are reviewed and evaluated data and information synthesized by a consensus of experts and periodically reexamined and updated. Another type of valuable research is the current critical appraisal of the methodological subject headings in MeSH (83). More explorations into the area of quality filtering would continue NLM's leadership role as well as benefit the field of biomedical communication and the health of the country.

The diversity of demands by specialized groups will continue to strain NLM's ability to acquire and organize needed scientific literature in a manner acceptable to all users, and comments on system limitations can be expected to persist. Balanced against such comments should be an acknowledgment of the success of MEDLARS—as measured by many factors, including its continued wide use and the vending of many of its data bases in the offerings of commercial information services.

Chapter 4

Private Sector Health Information Systems

Chapter 4

Private Sector Health Information Systems

INTRODUCTION

The growth and accomplishments of the private sector information industry in the past two decades are dramatic. In the area of information systems, there has been a tremendous increase in computerized data bases and on-line searching capabilities. Health-related bibliographic information is now available from not-for-profit institutions and for-profit corporations that are either devoted solely to information activities or operating as a division of a diversified conglomerate.

Some private sector organizations index and abstract biomedical and other health-related literature and produce printed indexes and computerized data bases. Other private sector firms, known as information services or vendors, sell on-line access to data bases usually leased from a variety of data-base producers in the public and private sectors. Still other firms—Bibliographic Retrieval Services and the Institute for Scientific Information—provide on-line searching of their own data bases, just as the National Library of Medicine (NLM or the Library) does for the data bases it produces.

When data bases are leased, vendors often add information to them, or provide a unique language of commands (software) that makes data bases from several producers compatible. Thus, through a single vendor, a user can efficiently retrieve information from a variety of bases.

This chapter describes the various bibliographic health-related information products that are available from the private sector in order to provide background information for the discussion of domestic policy issues in chapter 6. Owing to the nature of the congressional requests for this OTA study, system issues such as quality of private sector data bases and effectiveness of services are not considered.

HISTORICAL ORIGINS

The historical origins of the private sector abstracting and indexing services, often referred to as secondary services, are rooted in the professional or learned societies that flourished in the 19th century. Although these services began with manually produced print products, the majority today produce a machine-readable product or data base as well. Many of the data base producers in the scientific and biomedical field remained in the not-for-profit sector, others were purchased by for-profit firms, and still others were initially profitmaking enterprises.

In almost all cases, public support in varying degrees was made available to the organizations —either to assist in the planning for conversion from manual to machine production of their products, or to assist in the development of new, large data bases. For example, BIOSIS (formerly Bio-Sciences Information Service) and Chemical Abstracts Service both received support from the National Science Foundation, and Excerpta Medica received grants and contracts from the Public Health Service.

Private on-line services originated later than many of the private scientific and biomedical data bases, but also are an example of Government and private sector interaction. Much of the original development of information search services (vendors) as they exist today was sponsored in a not-for-profit environment, predominantly through the Government's research and development expenditures. Usually, the services had been fully tried and tested before they were incorporated into private sector products and services.

Indeed, two of the three major commercial bibliographic services in the United States, DIALOG Information Services, Inc. (DIALOG) and System

Development Corp. (SDC), were recipients of Government contracts. Lockheed Missiles & Space Co. (formerly DIALOG's parent company) "developed its commercial activity on a foundation of Government service contracts from the Office of Education (ERIC), National Technical Information Service (NTIS), and the National Agricultural Library" (18). In return, the Government received an expanded work force and needed expertise. The third major commercial bibliographic service, Bibliographic Retrieval Services (BRS), was initially part of a State university system. Thus, the public sector, as well as private capital and individual initiative, was instrumental in stimulating the extraordinary growth of the information industry.

DATA BASES RELEVANT TO HEALTH SCIENCES

A compendium of data bases, including those in life sciences, chemistry, and social sciences relevant to biomedical research or health services can be found in various directories, such as *Computer-Readable Data Bases: A Directory and Sourcebook* (167). Certain data bases—BIOSIS PREVIEWS, EXCERPTA MEDICA, IRL LIFE SCIENCES COLLECTION, and SCISEARCH, whose contents focus on the basic biological sciences and clinical medicine—are described in this section.

BIOSIS PREVIEWS

BIOSIS PREVIEWS is a large international bibliographic data base covering the most frequently consulted journals in the life sciences, and is therefore of significant relevance to biomedical research and clinical medicine. It is produced by BIOSIS, a not-for-profit organization founded in 1926 to serve the information needs of the life sciences community. The journals BIOSIS PREVIEWS covers focus on basic research in the life sciences.

BIOSIS PREVIEWS is produced in connection with and corresponds in coverage to the printed versions of *Biological Abstracts (BA)* and *Biological Abstracts/RRM (BA/RRM)*. *BA*, the larger, is an important bibliographic indexing and abstracting source covering primary research journals in the life sciences; *BA/RRM* is an analogous service covering reports, reviews, meetings, and books. BIOSIS PREVIEWS provides references to research literature in the life sciences, including agriculture, biochemistry, bioengineering, biophysics, ecology, experimental medicine, microbiology, and pharmacology.

Essentially all of the journal literature of original research in the life sciences is included. In 1981, 9,143 scientific journals from 116 countries were screened for inclusion, as well as books, monographs, meetings and conference proceedings, semipopular journals, research communications, and symposia. Journals comprised solely of articles on the life sciences are covered completely: every article and review is abstracted and indexed. Journals covering other sciences—physics, chemistry, and related topics—are screened for articles relevant to life sciences, and these articles are indexed and abstracted. Each article referenced counts as one citation (or "record").

Approximately 3 million records dating from 1969 are available on BIOSIS PREVIEWS. In 1982, BIOSIS will add 14,584 records from *BA* and 16,668 records from *BA/RRM* per month. *BA* and *BA/RRM* are updated twice a month. In 1981, more than 300,000 records were added; 315,000 will be added by the end of 1982. The citations and/or abstracts are keyed directly on computer tapes, which are mailed to vendors and are available to users approximately 5 weeks before the printed versions of *BA* and *BA/RRM* are available. Since December 1, 1980, abstract text from July 1, 1976, to the present is included to accompany BIOSIS PREVIEWS; these consist of English language summaries from *BA*. As of April 1982, access to BIOSIS PREVIEWS is provided by three commercial on-line search services in the United States and a total of 10 on-line services worldwide. There are also three off-line search services in the United States and 11 outside the United States with access to BIOSIS PREVIEWS.

BIOSIS is operated by a board consisting of the President of BIOSIS and 12 rotating trustees who serve maximum terms of 6 years. Board members are chosen to represent the Federation of American Societies for Experimental Biology, the American Institute of Biological Sciences, the American Association for the Advancement of Science, the National Academy of Sciences, and a changing group of disciplinary societies.

In 1982, BIOSIS charges commercial search services $5,000 per year (plus the cost of 48 tape reels at $17.50 each) for the current file of BIOSIS PREVIEWS, and $6,600 per year (plus the cost of 24 tape reels at $17.50 each) for the abstract text package. In addition, BIOSIS charges the services $20.00 per connect hour in on-line usage royalties. Commercial search services are charged for off-line prints of BIOSIS PREVIEWS only if their charges exceed $0.10 per citation. They are also charged $0.09 for each abstract text printed off-line. Commercial search services include BIOSIS royalty and printing charges and their online fees in their user charges.

EXCERPTA MEDICA

Excerpta Medica is an information retrieval service that provides descriptive indexes of biomedical and clinical literature. Sixty percent of its records include abstracts of the primary literature. Excerpta Medica was begun in 1946 as an independent not-for-profit foundation with some grant support from the U.S. and Netherlands' Governments. In 1971, Excerpta Medica was acquired by a private Dutch company, Elsevier-NDU, a major scientific publisher. Excerpta Medica now publishes two indexes of bibliographic references to the drug literature, 43 journals containing abstracts of articles published in the primary journal literature (see table 6), and 6 "core" journals containing abstracts of the most recent literature in discrete clinical specialties from the 43 specialty abstract journals.

A combination of professional, part-time, and volunteer abstractors and indexers—all practicing physicians—screen approximately 400,000 articles from over 3,500 journals each year to produce EXCERPTA MEDICA (or EMBASE), a computerized data base of references to 240,000 articles per year; of these, 150,000 are abstracted for the 43 specialty journals. The data base is updated weekly.

EXCERPTA MEDICA was designed by and is addressed to clinicians, as well as to research scientists. All indexing is done by practicing physicians who also select, approve, modify, and in some cases rewrite abstracts of articles. By screening 3,500 journals and reports of scientific symposia worldwide, EXCERPTA MEDICA covers

Table 6.—Excerpta Medica Abstract Journals

Anatomy, Anthropology, Embryology, and Histology
Anesthesiology
Arthritis and Rheumatism
Biophysics, Bio-engineering, and Medical Instrumentation
Cancer
Cardiovascular Diseases and Cardiovascular Surgery
Chest Diseases, Thoracic Surgery, and Tuberculosis
Clinical Biochemistry
Dermatology and Venereology
Developmental Biology and Teratology
Drug Dependence
Endocrinology
Environmental Health and Pollution Control
Epilepsy
Forensic Science Abstracts
Gastroenterology
General Pathology and Pathological Anatomy
Gerontology and Geriatrics
Health Economics and Hospital Management
Hematology
Human Genetics
Immunology, Serology, and Transplantation
Internal Medicine
Leprosy and Related Subjects
Microbiology: Bacteriology, Mycology, and Parasitology
Neurology and Neurosurgery
Nuclear Medicine
Obstetrics and Gynecology
Occupational Health and Industrial Medicine
Opthalmology
Orthopedic Surgery
Oto-, Rhino-, Laryngology
Pediatrics and Pediatric Surgery
Pharmacology and Toxicology
Physiology
Plastic Surgery
Psychiatry
Public Health, Social Medicine, and Hygiene
Radiology
Rehabilitation and Physical Medicine
Surgery
Urology and Nephrology
Virology

SOURCE: *Excerpta Medica*, 1982.

the foreign, especially European, literature very thoroughly, and provides English abstracts. The *Drug Literature Index and Adverse Reactions Titles* are considered to be especially thorough; for drug-oriented searches of the literature, Excerpta Medica's DRUGDOC provides very comprehensive and complete bibliographies. These are deemed to be especially useful in their coverage of drug testing conducted in countries other than the United States.

Each of Excerpta Medica's abstract journals is produced under the supervision of one or more section editors who are practicing medical specialists in the Amsterdam area. Each section also has an International Editorial Board that does not meet, but whose members are said to be available to provide advice on problems that occasionally arise in journal selection, classification, and terminology. A Board of Chief Editors coordinates the work of the section editors, and acts as liaison with the two executive chief editors, who are responsible for organization and management.

Two million records have been entered in EXCERPTA MEDICA since 1967. The data base is vended through DIALOG on a user fee basis, i.e., Excerpta Medica charges DIALOG a royalty each time a DIALOG customer accesses EXCERPTA MEDICA. DIALOG offers worldwide access to EXCERPTA MEDICA records entered since 1975, and is responsible for setting user fees.

IRL LIFE SCIENCES COLLECTION

Information Retrieval, Ltd. (IRL) is an independent, privately owned British company providing 17 abstracting services covering discrete areas in biological and medical sciences (see table 7). It began in 1966, with the publication of *Microbiology Abstracts,* and has grown steadily. IRL prides itself on complete and timely coverage of journals in its specialty areas, and on inclusion of books, conference proceedings, reports, patents, and the "fugitive" or "grey" literature (advertisements, announcements, unpublished and privately published reports). IRL currently screens 5,000 periodicals; its data base contains approximately 440,000 records entered since January 1978. The average growth has been 110,000 records per year, and the data base is updated weekly.

Table 7.—IRL Abstracts

- Animal Behavior
- Biochemistry Abstracts
  - Part 1: Biological Membranes
  - Part 2: Nucleic Acids
  - Part 3: Amino-Acids, Peptides, and Proteins
- Calcified Tissue Abstracts
- Chemoreception Abstracts
- Ecology Abstracts
- Entomology Abstracts
- Genetics Abstracts
- Immunology Abstracts
- Microbiology Abstracts
  - Section A: Industrial and Applied Microbiology
  - Section B: Bacteriology
  - Section C: Algology, Mycology, and Protozoology
- Toxicology Abstracts
- Virology Abstracts

SOURCE: Information Retrieval, Ltd., 1981.

IRL products provide global coverage (less than 37 percent of its sources are of U.S. origin) and full abstracts in English or foreign language sources. The quality of indexing is controlled by having all indexing and abstracting done in-house by university graduates in the life sciences. The IRL LIFE SCIENCES COLLECTION is available on-line in the United States and worldwide through DIALOG.

SCISEARCH

SCISEARCH is a multidisciplinary bibliographic data base prepared by the Institute for Scientific Information (ISI), a private for-profit company located in Philadelphia. ISI markets an assortment of services and publications designed to provide a comprehensive and coordinated information discovery and retrieval system. These include a weekly early alert journal (*Current Con-*

tents); a selective information dissemination service (*Automatic Subject Citation Alert*); and a document dissemination service (Original Article Text Service). It also produces three comprehensive on-line interdisciplinary data bases: *Science Citation Index (SCI)* (available on-line as SCISEARCH); *Social Sciences Citation Index* (available on-line as SOCIAL SCISEARCH), and *Arts and Humanities Citation Index* (also available in on-line).

SCI consists of four separate but related indexes: *Citation Index, Source Index, Permuterm Subject Index,* and *Corporate Index. Citation Index* is arranged alphabetically by cited author. A reference entry contains the primary author's name, the year the cited item was published, and an abbreviated name of the publication with its volume and page number. The source items citing a particular reference are arranged alphabetically by source author immediately under each reference item. *Source Index* is similar to an author index with the full citations—author, titles of article and journal, date of issue—to all source references. *Permuterm Subject Index* is an index in which all significant title words of the source items are permuted, forming all possible pairs of terms. Primary terms are arranged alphabetically, and all the co-terms occurring with a particular primary term are listed alphabetically under that term. *Corporate Index* lists source authors by both geographic location and organizational affiliation. *SCI* and SCISEARCH identify articles but do not provide abstracts.

SCI's coverage of over 3,000 journals is based on a principle which is different from that of other scientific retrieval services: the concept of citation indexing. Citation searching assumes that if a particular article or set of articles or author has material directly relevant to the subject of interest, then other articles or authors citing that material are also likely to be relevant. (This is also known as "forward" searching. (*Shepard's Citations,* a commonly used legal reference, uses this principle.) A *citation index* is a structured list of all the citations in a collection of documents.

Citation indexing takes advantage of the built-in linkages between documents provided by author's citations by listing together all items with common citations, and thus provides multidisciplinary searching capabilities. Thus, SCISEARCH is complementary to data bases built along traditional disciplinary lines.

SCISEARCH can trace a scientific idea forward in time, i.e., from an earlier cited article to a later citing article. It is based solely on the scientist to author's decisions to reference other articles. Since the intellectual intervention of human abstractors or indexers is not required, SCISEARCH reflects scientists' research practices and their own vocabularies. Because *SCI* is based on citation indexing, significant items of information other than journal articles can also be retrieved. They include technical reports, patents, and material published in books or the popular press—in fact, any item that is cited by the author. This citation feature also alerts users to relevant materials, including historical articles, that may have been written long before *SCI* was produced, and before other (alternative) data bases were accessible via computerized searches.

SCISEARCH contains approximately 3 million records and is updated biweekly. It is available on-line from 1974 to the present through DIALOG. It is priced differently to and by each search service, depending on the year accessed and on whether or not the on-line user is a subscriber to the printed version of *SCI*.

COMMERCIAL SEARCH SERVICES

In the private sector, the major on-line biomedical and health information search services are BRS, DIALOG, and SDC. They acquire data bases from a number of sources: private sector producers, both profit and nonprofit, and Government organizations, such as NLM, NTIS, and the National Agricultural Library, and offer their users the ability to search them with a single software system. Table 8 is a listing of prices to users for selected data bases.

Table 8.—On-Line Connect Hour Rates[a] for Selected Health-Related Data Bases, Spring 1982

Data base	BRS	DIALOG	NLM	SDC
BIOSIS PREVIEWS	$41-55	$43-58		$65
CA SEARCH	40-54	49-64		68
EXCERPTA MEDICA		55-70		
IRL LIFE SCIENCES COLLECTION		30-54		
MEDLINE	19	20-35	$15-22	
SCISEARCH		50-65[b] 150-165[c]		
HEALTH	15	20-35	15-22	
ENVIROLINE	61-75	63-78		90
TOXLINE			45-52	
CHEMLINE			94-101	
AGRICOLA	18-32	20-35		40
NTIS	24-38	25-40		45

[a] Prices are not directly comparable. Some information services may vend the same data base but the data base may span different periods of time. The telecommunications costs also vary, as does the office printing and other costs. For more specific information see tables 9, 10, and 11.
[b] Subscribers.
[c] Nonsubscribers.

SOURCE: BRS, DIALOG, NLM, SDC, 1982.

Each of the vendors offers a different package of software capabilities to its users. All major capabilities offered by the NLM MEDLARS II system are duplicated in the commercial search services. However, because of the intense competition in the information industry, the search services revise and expand their capabilities more rapidly than NLM. Their users can, for example, search whole data bases on phrases they compose themselves by asking for words to be in a certain proximity to one another. They can review their search strategies on the holdings of all or any group of data bases from one search service without having to reenter terms in one data base after another. They can order documents retrieved in on-line searches. They can put their own private files on-line for searching on the vendor's software. As noted in chapter 2, the plans for MEDLARS III include the acquisition of functions now available on commercial search services, except for the purely commercial ones, such as putting a user's private files on-line for searching. Samples of the capabilities specific to each search service are included in the following descriptions.

Bibliographic Retrieval Services, Latham, N.Y.

BRS initially focused specifically on the biomedical data bases, but it has since broadened its coverage by adding ones in science, business, and technology. The company now offers over 50 files on-line.

BRS began as the Biomedical Communications Network (BCN) at the State University of New York (Albany). BCN provided on-line access to nine data bases, including MEDLINE and BIOSIS PREVIEWS, to large universities and medical schools primarily in the Northeast. State funding for the project was withdrawn in 1975, and BRS was formed as a not-for-profit concern the following year, It was incorporated in May 1976 as a for-profit corporation. In 1980, BRS was purchased by Indian Head, Inc., a subsidiary of the Dutch company, Thyssen-Bornemisza.

User Profile

In the biomedical area, BRS estimates indicate that its users can be categorized by type of institution: 55 percent academic; 18 percent corporate industrial; 15 percent governmental (including the U.S. National Institutes of Health and 30 Veterans Administration hospitals); 6 percent miscellaneous (including public and State libraries). In terms of individuals, BRS estimates that 90 to 95 percent of its activity is handled by trained intermediaries (librarians or searchers) and that the remaining 5 to 10 percent is done by academic investigators in science. For 1980, BRS estimates that physicians comprised no more than 1 to 2 percent of its trained searchers. The number of users is considered proprietary information. Passwords are issued to both individuals and organizations. Health-related data bases offered by BRS are listed in table 9.

Costs

The BRS rate for searching the MEDLINE data base is $15 per hour, excluding telecommunications charges and the NLM $4 per hour royalty charge. The costs for accessing all other BRS data bases are shown in table 9. BRS' on-line connect hour rates depend on the number of hours of searching in 1 year. Users subscribe according to their projected use at the following rates. Any royalties which are charged by the data base producer are added to these subscription fees.

Table 9.—BRS: Selected Health-Related Data Bases Available On-Line, March 1982

Data base	On-line connect time rate per hour[a]		Off-line print rate per citation
	Minimum	Maximum	
Major relevance health care			
BIOSIS PREVIEWS	$41	$55	—
CA SEARCH	40	54	$0.10
HEALTH	15	15	—
MEDLINE	19	19	0.10
MEDOC	21	35	—
PRE-MED	16	30	—
Relevant to some specialties and interests			
AGRICOLA (CAIN)	18	32	—
DISSERTATION ABSTRACTS	24	45	
DRUG INFO AND ALCOHOL USE/ABUSE	21	35	0.05
ENVIROLINE	61	75	0.10
NRIC	21	35	0.05
NIMH (NCMH)	16	30	—
NTIS	24	38	—
POLLUTION ABSTRACTS	51	65	0.10
PSYCHOLOGICAL ABSTRACTS	46	60	—
SSIE	46	60	0.10

[a] Connect time costs include royalties paid to data base suppliers, but do not include telecommunications costs of $5 to $7 per connect hour.

SOURCE: Bibliographic Retrieval Services, 1982.

BRS On-Line Rates, February 1982

Number of annual connect hours	Annual subscription payment	Cost per computer connect hour
25	$ 750	$30
60	1,500	25
120	2,400	20
240	3,800	16

User Training

New users are introduced to the entire BRS system in a 1-day training session, which costs $35 for subscribers and $55 for nonsubscribers. BRS has a special 1-day training session for MEDLARS, which costs $50, and similar sessions for other data bases. Approximately 300 training sessions are conducted each year. The sessions are held in major U.S. metropolitan areas on a monthly or bimonthly basis, and in other onsite locations at the request of users or potential customers.

Value-Added Services

BRS has also created a file called PRE-MED that updates the primary English-language medical journals weekly with bibliographic citations not yet available on MEDLINE. BRS also carries the full text of all 16 of the journals from the American Chemical Society.

BRS places a particular emphasis on marketing a private-file service. Users can mail them machine-readable files of their own data—library catalogs, academic departmental publications lists, etc.—and BRS will put the data on its computers and offer on-line retrieval with its software at a competitive price. If the information is not currently machine-readable, BRS offers a number of options from a preprogramed microcomputer for local input to a direct on-line data input program. BRS also offers electronic messaging for communication among groups of users and an electronic newsletter capability.

DIALOG Information Services, Inc., Palo Alto, Calif.

In 1966, DIALOG Information Services (DIALOG), then a subsidiary of the Lockheed Missiles & Space Co., developed the first on-line retrieval system that went beyond the demonstration stage to regular production (18). This system, also called DIALOG, is the basis for systems currently in operation at several Government agencies, including the National Aeronautics and Space Administration and the Justice Department. The system has been commercially available since

1972, and now provides access to over 120 data bases for 10,000 users in 40 countries (18). Six of its data bases are of major relevance to health and biomedicine.

Initially, DIALOG focused on education and science, but it now offers data bases covering business, social science, and the humanities. The company is expanding DIALOG's health coverage: INTERNATIONAL PHARMACEUTICAL ABSTRACTS and EXCERPTA MEDICA were added in 1975; MEDLINE was added in 1981; and NLM's TOXLINE and HEALTH data bases are expected to be added in the near future. This firm is the largest commercial vendor of on-line biomedical bibliographic information in the world.

User Profile

The DIALOG system has over 10,000 users. It is company policy not to disclose specific information about users because of competition with other vendors. "Users" in this case means "passwords" and is roughly equivalent to organizations. However, some complex organizations and institutions may have several libraries or information centers with terminals and thus have several passwords.

Costs

DIALOG's access charges are displayed in table 10. A number of discounts are available, all based on monthly usage levels. The maximum discount of $15 per hour is available for a guaranteed usage of 80 hours per month.

Like BRS, DIALOG bases its prices at least in part, on a projected market value of each data base and service. Market value is determined by the nature of the data base, the user population (e.g., financial communities are likely to pay more for a service than academic users), and competitors' prices.

User Training

DIALOG makes its services available to potential users by conducting 1,200 training sessions in 100 U.S. metropolitan areas and other sites at customer request. These consist of a general 1½-day training session for use of the entire DIALOG system, augmented by specialized training sessions for a half-day each on EXCERPTA MEDICA, on MEDLINE, and on BIOSIS. The charge is $65 for the general training session (which includes on-line practice) and $25 for the specialized sessions. These courses are attended mostly by librarians or specialized on-line searchers in universities, and in increasing numbers by physicians and scientists (end users).

All DIALOG data bases are completely on-line. Unlike NLM or BRS, DIALOG does not put earlier years of data bases into backfiles for off-line processing; DIALOG puts them into backfiles for on-line processing. DIALOG also offers a private file service. The company's DIALORDER system covers more on-line ordering sources than any other vendor. The company also carries electronic newsletters and on-line directories.

System Development Corp., Santa Monica, Calif.

SDC was established as a not-for-profit company in 1956 as part of the Rand Corp., a consulting firm established after World War II to provide research services to the Air Force. SDC initially trained operators and programers for the SAGE early warning system. SDC became a freestanding, for-profit corporation in 1968. It has been involved with biomedical and health services data bases since 1968, when it assisted NLM in developing ELHILL, the software package for MEDLARS. SDC modified the ELHILL software for the ORBIT package it now uses. The company was purchased by the Burroughs Corp. in 1980. It has more than 80 data bases that provide coverage of information in many areas of knowledge.

User Profile

There is no systematic information available about SDC users in the biomedical/health services area, but an estimated 15 to 20 percent of its total business is in the medical area. Of SDC's total customer population, an estimated 60 to 65 percent are from private business and industry, 15 to 20 percent from Government agencies, and the remaining 15 to 25 percent from academic institu-

Table 10.—DIALOG: Selected Health-Related Data Bases Available On-Line, April 1982

Data base	On-line connect time rate per hour[a]		Off-line print rate per full report
	Minimum[b]	Maximum	
Major relevance to health care			
BIOSIS PREVIEWS (1969 to present)	$43	$58	$0.15
CA SEARCH (1967 to present)	49	64	0.20
EXCERPTA MEDICA (1974 to present)	55	70	0.20
HEALTH	20	35	0.15
IRL LIFE SCIENCES COLLECTION	30	45	0.15
MEDLINE (1966 to present)	20	35	0.15
SCISEARCH (1970 to present)			
(subscribers)	50	65	0.15
(nonsubscribers)	150	165	0.25
Relevant to some specialties and interests			
AGRICOLA (1979 to present)	20	35	0.10
CHEMNAME	115	130	0.20
CHILD ABUSE AND NEGLECT	20	35	0.10
COMPREHENSIVE DISSERTATION INDEX	40	55	0.12
ENVIROLINE	63	78	0.15
ENVIRONMENT BIBLIOGRAPHY	45	60	0.15
FOODS ADLIBRA	40	55	0.10
FOODS ADLIBRA	40	55	0.10
INTERNATIONAL PHARMACEUTICAL ABSTRACTS	35	50	0.15
NTIS	25	40	0.10
PHARMACEUTICAL NEWS INDEX	80	95	0.30
POLLUTION ABSTRACTS	58	73	0.20
POPULATION BIBLIOGRAPHY	40	55	0.10
PSYCHOINFO	50	65	0.10

[a] Connect time rates include royalties paid to data base suppliers, but do not include telecommunications costs of $6 per connect hour.

[b] Minimum connect time rate applies for users with a discount contract.

SOURCE: DIALOG Information Services, Inc., 1982.

tions. At the present time, SDC is persuaded that the majority of its users in the biomedical area are from pharmaceutical firms, with very few from academic or medical institutions. SDC hopes to attract a larger portion of the health and biomedical consumer market when it adds MEDLINE and TOXLINE, partly in response to requests from customers and reports from field representatives.

Costs

Three of the SDC offerings have annual subscriptions for the printed indexes produced from each data base, but SDC has no subscription costs, startup fees, or monthly minimums of its own. Discounts are available to users for more than 5 hours of connect time each month, to a maximum discount of $20 for 140 hours per month. SDC's charges are shown in table 11.

User Training

SDC holds regular training sessions for new users in major cities. Approximately 125 sessions are held each year in 18 locations. Users are taught the basic techniques of the ORBIT software system by professional instructors. New-user training sessions last 1½-days and cost $150 per person; this fee is credited back to the account of the user. The session includes a training package, online practice, and 3 hours of computer time. There are additional half-day seminar sessions for life sciences data bases, at $50 per session. In addition, custom-designed classes are available at the

Table 11.—SDC: Selected Health-Related Data Bases Available On-Line, March 1982

Data base	On-line connect time rate per hour[a]	Off-line print rate per citation
Major relevance to health care		
BIOSIS PREVIEWS (1969 to present)	$65[b]	$0.10-0.25
RINGDOC[c] (1964 to present)	100	0.13
CA SEARCH (1967 to present)	68	0.20
Relevant to some specialties and interests		
AGRICOLA (1978 to present)	40	0.06
CHEMDEX	125	0.25
ENVIROLINE	90	0.20
NTIS	45	0.10
PESTDOC[c]	100	0.13
PSYCINFO	65	0.10
SAFETY SCIENCE ABSTRACTS	75	0.15
SPORT	85	0.15
SSIE	110	0.25
VETDOC[c]	100	0.13

[a] Connect time costs include royalties paid to data base suppliers, but do not include telecommunications costs of $8/hour.

[b] SDC also has a discount plan that applies to all accounts that use SDC's system, ORBIT, for at least 5 hours in a given month. The discount is applied in steps dependent upon the number of connect hours used in a given month on all data bases.

[c] RINGDOC, PESTDOC, and VETDOC require a yearly subscription to the printed journal from Derwent Publications before they can be accessed on SDC. These subscriptions costs as of 1981 were:

PESTDOC	$11,275/yr
RINGDOC	30,350/yr
VETDOC	7,625/yr

SOURCE: Systems Development Corp., 1982.

customer's location on request, at a cost of $450 plus the trainer's travel costs.

Value-Added Services

SDC has recently developed advanced cross-file searching capabilities for its ORBIT software comparable to that used by DIALOG and BRS. Users can create their own thesaurus of terms by browsing through lists of terms in one data base and carrying those lists over into other bases. If a searcher finds a particularly interesting citation in one file, he or she can command the system to enter the terms appearing in that citation into their search strategy automatically. Searchers more familiar with softwares other than ORBIT can rename commands and have those changes automatically executed whenever they enter their password. SDC also developed the first on-line ordering system, the Electronic Maildrop, whereby users can place an order for an original document at their terminal.

Chapter 5

Public and Private Information Sectors: Elements of Current Domestic Policy Issues

Chapter 5

Public and Private Information Sectors: Elements of Current Domestic Policy Issues

INTRODUCTION

The Federal Government has had a longstanding role in the creation and distribution of information goods and services. At the same time, the private sector has participated extensively in Government information matters. For example, all through American history, the Government has had a significant role in publishing, yet the main publishing activities in this country have been private. The role of the Government has been neither to overwhelm or to restrict, but rather to encourage private information activities. Interactions between the private and public sectors are and have been dynamic but often discordant, requiring continuous renegotiation. In recent years, the growing reliance of society on information and the increase in the number and type of information products, services, and resources have heightened the tension between the two sectors.

There are fundamental philosophical differences underlying the discord. One philosophy is that the Government's role should be restricted in information activities and that information is best managed and distributed in the marketplace by having many non-Government information sources. The other philosophy recognizes the economic value of information, but in addition is concerned with its social value and impact (125). As a result of the philosophical differences, agreement as to the proper role of Government has proven elusive.

After 2 years of deliberations, a task force of the National Commission on Libraries and Information Sciences (NCLIS) completed a thoughtful study setting forth *general* principles concerning the role of the public and private sectors in Government information activities. The NCLIS task force is "in favor of open access to information generated by the Federal Government; in favor of reliance on the libraries and private sector organizations (both for-profit and not-for-profit) to make readily available information that can be distributed by the Federal Government; in favor of a leadership role for Government rather than a management role; and in favor of limiting direct Government intervention in the marketplace" (96). (See app. F for a complete listing of principles and recommendations of the NCLIS Public Sector/Private Sector Task Force.)

This chapter presents primary considerations for examining future as well as present national information issues and the specific issues concerning the National Library of Medicine's (NLM or the Library) creation and provision of computerized bibliographic information. The considerations are the Government's role in the allocation of resources to information development and distribution, and the effect on the private information sector of the Government's involvement in allocative activities. The influence of the Government's pricing policies is of particular importance. Because a historical perspective is needed to understand Government information activities, the chapter includes a brief section on the history of public and private sector involvement in health information policy.

DEFINITIONS

Except where explicitly stated otherwise, OTA uses the definitions of the NCLIS Public Sector/Private Sector Task Force (96):

> Public Sector—This term . . . includes Government and, more specifically, Federal Government. Agencies, like public libraries or public universities that are entirely tax-supported, even though non-Governmental in character, are included.
>
> Private Sector—This term . . . includes private enterprises, for-profit and not-for-profit, as well as organizations such as professional societies and trade associations, hybrids that are joint Government/private enterprises, and organizations such as privately supported libraries and universities (even though they may be subsidized by public funds).

HISTORICAL PERSPECTIVE

Health information has a somewhat different policy history than that of other science and technology information. Studies and activities in the late 1950's and early 1960's that dealt with improving scientific and technical information either implicitly or explicitly treated health information in a different manner from other information both in their general considerations and when addressing public/private sector issues. After the mid-1960's, most information policy studies and activities, including those dealing with public/private sector issues, were not discipline oriented, and, thus, did not consider science information by itself.

In addition to being allied to information policy, health information policy has been and is now strongly tied to health policy, particularly that of biomedical research and medical education. In the past decade, health information policy has come to be increasingly associated with policy toward health services delivery, behavioral research, and Federal payment for biomedical, behavioral, and health services research.

Many Federal actions and publicly sponsored studies of health information communication or health issues testify to the perception that health information is singular, and requires public support to assure its production and dissemination in the interest of the Nation's health. A comprehensive review of the roles of the public and private sectors in health information policy is beyond the scope of this report: the following selected studies and legislative and executive actions exemplify the public position.

Perhaps the strongest indications of the attitude of both the executive and congressional branches of the Government toward health information are the passage of the National Library of Medicine Act in 1956 (Public Law 84-941), which established NLM, and the continued appropriation of funds for the Library's operations. (See app. A for a discussion of the legislation.) The National Library of Medicine Act was passed without opposition. Witnesses from the Government and the medical profession unanimously supported it as a most important contribution to biomedical research and the national welfare. The White House was convinced that the Library would serve as the best resource for the production and dissemination of biomedical information (13). NLM was placed under the auspices of the Public Health Service both to foster contact with programs in biomedical research and to assist the Library with providing access to information about scientific advances to the health community.

NLM began to assume its official role as a national resource at the start of a period of governmental examination of the importance of research and development (R&D) in the sciences. The 1958 Baker report (122) was the first of a series of studies of information handling that arose from concern with American science and technology after the launching of the Soviet satellite Sputnik I in 1957. It implicitly separated health from other scientific information in its recommendations. One result of the report was the establishment of the Office of Science Information Service in the National Science Foundation. The office was given congressional and Presidential authority to assume leadership in coordinating scientific (excluding medical) information activities within the Government and between the public and private sectors. For the most part, private information services at the time of the report were in the not-for-profit sector of the industry.

During this period, the Subcommittee on Reorganization and International Organization of the Senate Committee on Government Operations was deeply involved in studying the management of scientific and technical information, and from 1960 to 1962, the subcommittee released a number of documents on various aspects of the issue. As the major supporter of scientific and technical R&D, the Government was perceived as having the major responsibility for information. The little attention the subcommittee paid to the private sector consisted of urging scientific and technical professional and trade societies to "meet their own challenges head-on and not wait upon Government to do so" (147).

Although the subcommittee was very concerned with the transmission of all categories of scientific and technical information, it had a stronger interest in biomedical information. The stimulation of the not-for-profit organizations in the private sector was a small part of the subcommittee's interest, but interaction between the public and private sector was looked on favorably. Although the greatest emphasis was put on the Government's responsibility for biomedical communications, Senator Humphrey was interested in public/private cooperation and pointed to the publication of the annual *Cumulative Index* (which at that time was published by the American Medical Association based on NLM's monthly *Index Medicus)* as a fine example of collaboration between the two sectors (147).

In 1962, the Surgeon General's Conference on Health Communications emphasized the Public Health Service's responsibility in improving the communication of biomedical research (138). A major recommendation was that the Public Health Service should "give technical libraries support for their present activities and make funds available so they can use, experiment, and broaden their role in meeting the needs of users, including scientists, health practitioners, health educators, and science writers."

In 1962-63, at the request of the National Institutes of Health, the National Academy of Sciences/National Research Council conducted a study of biomedical research communications problems (111). The report recommended that the biomedical community continue managing its own communications. Although it conjectured that the growth in biomedical research would require public funding of some information modalities, it stressed the need for scientific control of the communications process. The report differentiated the conduct and location of biomedical research from other scientific and technical research, and suggested that the type of communication used for other types of scientific research might not be appropriate for biomedical research. It singled out NLM as "the central resource for the network of biomedical libraries and information services and as the major indexing service in the biomedical field . . . and the hub of the entire document retrieval component of the biomedical communication complex" and urged its continued support by the biomedical community in its future development. It also urged support in the form of direct grants in aid in the short term to upgrade local biomedical academic libraries.

A similar and even stronger position was taken by the President's Commission on Heart Disease, Cancer, and Stroke in its report of 1964 and its source papers of 1965 (121). The report dealt at length on the needs for improved health communications, and advised that this should be accomplished through existing libraries throughout the country. The report noted that as a result of the explosion in biomedical research "the continued and accelerated generation of scientific knowledge will become increasingly an exercise in futility" without an improvement in the library school base, as medical libraries were essential for advancing health knowledge, health education, and health practice. It emphasized the vital need for Government leadership in providing assistance to medical libraries to assist researchers and practitioners in fulfilling their information needs, and recommended strengthening NLM and bolstering the Nation's medical school libraries. It suggested establishing a library network based on existing libraries with centralized responsibility in the NLM, building on its acknowledged capabilities, including that of MEDLARS. This report provided a primary impetus for the passage of the Medical Library Assistance Act of 1965 (Public Law 89-241), which incorporated many of its recommendations. (See app. A for a discussion of the legislation.)

Further expressions of congressional and executive interest in public support of health information and NLM are found in the transfer of the Public Health Audiovisual Facility to NLM in 1967. The same year, the Toxicology Information program was assigned to NLM, and the next year, the Lister Hill National Center for Biomedical Communications was set up as part of NLM.

The uniqueness of the Government's involvement in health information was later underscored in a report on scientific and technical communication sponsored by the National Academy of Sciences and the National Academy of Engineering (95). One of the 1969 report's recommendations was that steps should be taken to upgrade and stimulate the initiation of privately operated information services, which would then serve as component elements in information programs with Government services. The report explicitly excluded NLM from this recommendation, thereby acknowledging the need for Government-funded health information services.

More recent studies of information policy deal with broad issues that overlap disciplines, such as privacy, Government management of data processing, and the need for a national information policy (see app. B). Thus, few studies base their considerations on health as distinct from other disciplines. Health policy studies that consider information issues and health information studies focus on such issues as quality filters for biomedical information, means of informing the public of biomedical advances, and means of increasing the efficiency of the Government's management of health information; they do not consider the Government's role in providing health information to any degree.

Government activities in health information have been concentrated mainly in the areas of establishing health-related clearinghouses and other types of information organizations, attempting to establish a coordinated Health Information System based on primary data, appropriating funds under the National Library of Medicine Act, and reauthorizing and appropriating funds for the Medical Library Assistance Act. At many of the hearings regarding these two acts, there was considerable questioning with respect to the Library's cost recovery practices for its information goods and services, but little attention was paid to the appropriate role of the Library in providing information goods and services.

Thus, public policy regarding the degree of Government involvement in health information activities in general, and in NLM in particular, has persisted essentially as enunciated 25 years ago. In the past few years, however, there have been definite indications that this policy is being reconsidered. The section of this chapter entitled "NLM's Pricing Policies" addresses the proposed policy changes.

THE APPROPRIATE ROLE OF GOVERNMENT IN INFORMATION ACTIVITIES: MAJOR UNDERLYING CONSIDERATIONS

Introduction

In this Nation, there is a clear preference that the private sector produce and offer goods and provide services. This is underscored in OMB Circular A-76 (1979) (49), which sets policies for the acquisition of industrial products and services needed by the Government. The circular states that the "Government should not compete with its citizens" and that it is "the *general* policy of the Government to rely on competitive private enterprise to supply the products and services it needs" (emphasis added).

The complexity of the issue is illustrated by the policy precepts which form the foundation of the *general* policy. As stated in OMB Circular A-76, the three "equally valid policy precepts" are: 1) to rely on the private sector, because "the Government's business is not to be in business;" 2) to retain certain Government functions in-house, because certain functions are "so intimately related to the public interest as to mandate performance by Federal employees;" and 3) to aim for economy and employ rigorous cost comparisons when deciding how work should be done.

Thus, drawing the line between those Government activities which supplement or promote private efforts and those which abridge private enterprise is difficult. Economic considerations are of some help in identifying appropriate Government activities. This section discusses the Government's role in the allocation of resources to information development and distribution, and the attendant consideration of the effects of the Government's involvement in information activities on the private information sector, particularly the effects of the pricing of Government products and services.

Allocation of Resources

Economic Theory

In general, the market mechanism for the allocation of resources has proven to be effective, and in the United States, most goods are produced and bought and sold in the marketplace. There are several reasons why Government itself assumes the functions of allocating resources and distributing certain goods and services. One reason lies in the nature of the goods and services: some goods and services are available to everyone if they are available at all. No one can be excluded from the service. Services such as national defense have the property of nonexclusivity. Such services can be provided more efficiently by nonmarketing techniques. Other examples are public health measures to reduce contagious diseases, such as eradication of the anopheles mosquito that carries malaria. Goods and services with this property of nonexclusivity are termed "pure public goods" by economists.

Another possible justification for governmental allocation is the presence of positive externalities—i.e., the total social benefit from a good or service may be greater than the sum of the benefits to separate individuals. For example, the social benefit from the research that resulted in the development of the polio vaccine was far greater than the sum of the benefits to the individual researchers and firms that participated in the experimentation or to the people who received the vaccine. The research yielded benefits external to those separate individuals. The society as a whole gained added benefit from the reduction of a major health problem. In such cases, Government funding of the research may be considered appropriate in order that society reap more of the potential benefits.

Although there are other economic reasons that justify the Government's provision and distribution of goods and services, the two just discussed are the most pertinent to the issues of this OTA study. The allocation of resources by the Government is based on both social and economic values. Regardless of the justification for its involvement, the Government may provide a good or service directly or it may purchase the good or service from the private or public sector.

Allocation of Resources to Information Development and Distribution

One reason it has been difficult to clearly define the role of the Government in information activities is that economic theory concerning the allocation of resources has significant limitations, although it provides many useful explanations. Economic theory does not always conform nicely with reality. Another reason is the lack of general agreement as to the nature of information. Information can be conceived of in two ways: some understand it to mean the content of communication, while others equate it with the medium of communication. This difference is seen in the characteristics that are attributed to information (96,12,29,21):

- Information is an intangible.
- Information can benefit an individual recipient and/or society as a whole; it is not exclusive.
- Information is not depleted by use; if one person uses it, this does not mean that another person will not be able to use it. The value

of the information may decline, however, even if the information itself does not change.

- Information can be possessed by more than one person at the same time.
- Information cannot be easily packaged in well-defined units.
- Information can be made available in many media, such as books, documents, journal articles, bibliographic references, videotapes. These forms have many of the same characteristics as other goods and services.
- Information is marketable and may be profitable. It has value as a capital resource, as an essential tool for decisionmaking, and as a means for the better management of tangible resources (118).
- The price of information bears little relationship to the costs of making copies available. The cost of the first copy is likely to be a very large part of the total cost of production and the reproduction costs are relatively minor (96).
- The value of information increases as the amount of data involved and as the degree of organization of those data increases (96).

When information is viewed as a medium or conduit, it does not possess the characteristic of nonexclusivity, the common characteristic of a pure public good. One reason for the Government to create or distribute information is that it is a public good. However, most Government information activities are conducted because of the presence of externalities. For example, when the Government provides information about impending typhoons or hurricanes, not only do the people in the path of the storm benefit from the information, but society as a whole benefits from the reduction or prevention of a disaster.

Externalities from a particular good or service are not always similarly perceived by all members of society. In the United States, the Government provides an extensive crop-reporting service that is invaluable for agricultural markets. The Government also gathers statistics forecasting total production (gross national product) which, in turn, are used by private firms and trade associations to forecast their sales in order to plan production schedules and major long-term investments. Some would contend that the Government is not justified in providing access to such information in either case because the benefits of the service to society at large are not greater than that of the sum of the benefits to separate individuals. Others would argue that such benefits, in the long run, are greater to the society as a whole than to the individuals.

The problem is complicated by the extreme variety in the content and medium of information. The responsibility for creating and disseminating information may rest with the Government or the private sector depending on its content and/or media. Because of considerations of public good or externalities, which are influenced by societal values, the Government may be justified in certain information activities.

On the other hand, the long history of the private sector in publishing activities and the growth and development of the commercial information field in the last two decades give evidence that much information can be successfully provided and allocated by the market. But private enterprise is, and of necessity must be, selective in the information it provides, in order to stay in business. The market may not operate successfully for all types of information, particularly highly technical and esoteric materials. Such materials may not have a large enough audience to be profitable for a commercial firm to market, although the information is or may be essential for the common good. They may or may not be of sufficient interest to nonprofit professional associations.

Effects of Government Involvement in Information Activities on the Private Information Sector

Introduction

Where one draws the line separating competitive from complementary activities of the public and private sectors depends on one's philosophical orientation. Some believe that certain Government activities amount to unfair competition with the private sector, while others do not. There are areas where the private and public sectors are in competition (e.g., both are involved in providing postal services, education, and weather information). In most cases, prices that the Government charges for these services are zero or very low. Some have interpreted this as unfair competition.

In the current dialog among the sectors involved in information activities, the term "unfair competition" is not used in the legal sense. The laws that govern competitive activities apply only to private actions, and the Federal and State Governments are deemed to be immune from their prohibitions. The term "unfair competition" is used in the sense that prices for Government-sponsored information products are often subsidized, resulting in unfair price competition.

Regardless of definitional distinctions, the presence of the Government in the information market does influence the private information sector. Because the two sectors are not subject to the same laws, conflicts between Government and private activities are accentuated.

Market Segmentation

In evaluating whether or not the Government is competing unfairly with the private sector, the boundaries of the market for the information good or service under consideration need to be defined. In general, products and services that differ compete in discrete markets. For the most part, the Government does not offer the same product or service as private industry. A Government-provided information product, for instance, may contain data that is not of interest to a limited market or may not be of immediate interest. In addition, the Government may make information goods or services available to populations or areas that private firms would not find profitable to serve. In some cases, however, Government information goods and services are similar to ones produced by the private sector.

Assessing whether or not information products and services are competing in different markets depends not only on the specific characteristics of the product or service, but also on the substitutability of one product or service for another in the market. For example, the mail, the telephone, and the telegram can be considered as substitutes for one another and as such are competing in the same market. In reality, the substitution may be imprecise, and these three communication modalities can be considered as operating in three distinct markets. Geographic distribution may be another factor in considering substitutability and market segmentation. The services provided by one telephone company may be similar to those provided by another. But if one telephone company is serving a more circumscribed area than another, the two companies can be considered as operating in discrete markets.

Segmentation may also occur when there are different types of buyers for the product or service. For example, the users of health-related information can be conveniently divided into students, teachers, investigators, practitioners of various categories, managers, administrators, planners and policy analysts, the Government, health-related industries, publishing and communications industries, recipients of health-related services, and the general public. The markets are quite discrete, and the buyers in different markets may be willing to pay different prices (65).

The Information Market

Information transfer is in an area where market and institutional imperfections exist (125). Factors leading to market inefficiencies include economies of scale, economies of scope, and structural and price barriers that prevent entry into a market (21). Information goods and services are often produced with significant economies of scale due either to extremely high startup costs or fixed costs or to the existence of decreasing unit costs—or both.

The situation is even more complex because information production and distribution also leads to economies of scope. Since doing so is often less expensive, most information organizations usually produce more than one product or offer more than one service: e.g., a telephone company usually offers long distance as well as local call service, or the publisher of a printed abstract journal may offer the same information in a computerized data base (21).

Structural and pricing barriers that may impede the entry of new firms into the field include the "lack of availability of major resources because of uniqueness, geography, and so forth; lack of competition because of low pricing and Government regulatory restrictions that exclude additional firms from certain markets; and market restrictions caused by the advantages of scale economies" (21). Other factors, including the difficulty of determining market demand and uncertainties about the value of information, may also bring about imperfections in information markets. Lave believes that as a result of the above factors, the private sector may tend to undersupply information, particularly information that is needed by and valuable to a small number of users (80).

Government and the Information Market

The Government both corrects and causes imperfections in the information market. The inefficiencies in the market provide a rationale for a Government role in information activities (80). Some of the inefficiencies in the market, however, are the result of Government participation.

The Government enters the market when the externalities are important enough that the market cannot provide an efficient outcome and the economies of scale required to produce or distribute a product or service are such that the Government is about the only organization that has sufficient resources to do so. The Government can also correct market imperfections by subsidizing the private sector to provide the information (80).

When the Government undertakes the production and distribution of information, it may have the power to affect the market price. Most Government organizations price their goods and services at zero or a low price. In addition, the Government often has the advantage in the marketplace of prestige. There is a perception on the part of some users that Government-sponsored information goods and services are reliable (168). Thus, because of the low price and/or perceived quality of Government-sponsored information goods and services, some buyers will favor these goods and services over those that are produced or disseminated by commercial firms. Hence, new entrants may be dissuaded from entering the information market, and if the Government enters an established market, others may leave.

Pricing of Information Products and Services

Introduction

When the Government participates in the development and provision of certain information products and services, the questions are whether, when, and what to charge users for such products and services. What rates should an agency charge to best serve the public purposes for which the agency was created? What rates should an agency charge that will not stifle the growth and development of the private information sector?

Although the library tradition is that ideally all information should be made freely available to all, some libraries have always charged for some services. Many first introduced fees with the advent of computer-based reference services (49). Most services allocated by the public sector are mixed with private benefits. It has been suggested that user fees be instituted when the following conditions obtain (92):

- Units of service can be defined and measured, and the fees related to benefits received in an equitable fashion.
- Individual users can be identified, and ownership rights defined: that is, at least some of the benefits are private.
- Fees can be enforced and nonpayers excluded at a reasonable cost.
- Charging users is not contrary to other, overriding social objectives: in particular, the distributional consequences of fees must be acceptable.

Practically, user reactions to fees must also be considered, and the following ideas for setting fees have been put forth (90):

- New fees should be for new services, not those which have previously been free.
- The fee should be closely associated with the service for which it is paid, and should be a simple function of the quantity of use, so that the consumer can make a rational decision about how much to spend.

- The highest amount consistent with the purpose of the public activity should be charged. As long as the institution is going to incur the political and economic costs of introducing fees, it is better to get full payment, not just the illusion of it.

Computer-based services fit most of the principles noted above. They differ from services traditionally offered in libraries, are relatively expensive, are tailored to individual needs, are obtained from an outside source (a vendor) who charges the library, and are easily identifiable and provided at the user's request.

Methods of Pricing Information Goods and Services

When the true costs of a good or service are known, average cost pricing or marginal cost pricing can be used as the pricing method. The costs of an information good or service are the generation of the information or the collection and organization of information; the design and development of the product or the service; the maintenance of the operation, including the equipment, rent, utilities and space (fixed costs); the handling of requests; the reproduction of materials; and the distribution of the good or service.

In the information field, average cost pricing is used to recover fixed costs and those associated with the handling of requests, the reproduction of materials and the distribution of the product and the service (72). Creation costs have not been included in calculating average costs for most information products and services. The creation costs of Government-sponsored computerized bibliographic data bases have been subsidized by the taxpayer. The creation costs of privately produced computerized bibliographic data bases have been in almost all cases subsidized by a companion print product. But because of usage and technological changes in the information market, some adjustment in cost calculations may have to be made for the increasing use of data bases and the decreasing use of print products.

When the true cost of a good or product is not available, formula pricing or target pricing is used. In formula pricing the manufacturing cost is multiplied by a predetermined factor. For example, books are often priced at five or six times the cost of printing. Target pricing is based on estimating the volume of demand and determining in advance a price that will yield a desired rate of return on the total cost (12).

Some information services charge different prices to different buyers for the same product or service for reasons not associated with costs. A problem with this pricing strategy—price discrimination—is the possibility that a buyer who bought the product at a lower price might resell it to others at a price lower than that charged by the first seller. Another pricing strategy not based solely on costs is pricing to achieve an objective and not necessarily to recover the costs of a particular product.

Pricing Practices of the Private Sector

Generally, the for-profit sector of the information industry depends on venture capital or profits from other products or services to initiate a new service or product (12). The price to the user of the product or service is not solely based on costs, but is usually established according to the perceived value of the product or the service to the user (66).

The information industry views itself as a "niche" industry in which each firm operates in a particular segment of the market. A firm carves out a portion of the market where it sees a need for its product or service and then develops the market for its specialized service or product (66).

The two current commercial vendors of NLM's data bases—DIALOG Information Services, Inc. (DIALOG) and Bibliographic Retrieval Services (BRS)—view themselves in this light. When BRS was established as a for-profit venture in 1976, it targeted the academic and medical library communities, high-volume users of biomedical data bases, as its market segment. The two principal organizers of BRS, Janet Egeland and Ron Quake, came from the (former) State University of New York (SUNY) Albany Biomedical Communications Network (BCN), a nonprofit agency. Using SUNY Albany computers, BCN served large universities and medical schools from the East Coast to Chicago during the late 1960's until 1975, when State funding for the project was withdrawn. At that time, BCN made available MEDLARS and eight other data bases: BIOSIS PREVIEWS, CHEMICAL ABSTRACTS, CAIN (now known as AGRICOLA from the National Agricultural Library), ERIC, INSPEC, NTIS, PSYCHOLOGICAL ABSTRACTS, and INFORM.

When first formed, BRS contacted members of its delineated market to determine the services they wanted and the price they would be willing to pay. The company then and now establishes prices on the market demand for the service and adjusts the service to the established price. Since its founding, BRS has added other types of data bases and expanded its market interests, particularly in the area of the corporate community (46).

DIALOG terms itself a service bureau industry specializing in bibliographic retrieval service for institutions. The company, which started as an internally funded R&D project of Lockheed Missiles Space Co. in 1961, made on-line services commercially available to the general public in 1972. It initially sold access to data bases in education and science, but now it includes data bases in many fields including health. DIALOG's pricing includes both cost and market value elements. Calculations of average cost are based on the following cost categories: royalties, telecommunication, equipment leasing, storage, loading and updating of data bases, documentation, facilities, personnel, and various overhead figures. DIALOG does not do rigorous market surveys before marketing a new data base, but has experience in estimating the market demand at various price levels for a new product.

In determining market value, the nature of the data base and its general potential of appeal to a mass market are taken into consideration. The segment of the market the data base will appeal to is also influential in establishing a price—e.g., financial market is more likely to accept a higher price than an academic community. DIALOG also examines what competitors are charging for a similar service. The value of the data bases to the mass market and/or market segment is balanced against the calculated cost, and a price determination is made after a profit is also factored in. However, if the market value appears to be lower than anticipated costs, the decision is made not to provide the service.

Government Pricing of Information Goods and Services: Federal Policy and Practices Currently in Force

The Federal Government has no set policy for pricing the information goods and services it provides. Authority for pricing stems from enabling legislation and the Office of Management and Budget's (OMB) directives, with departmental policies serving as a secondary source. Enabling legislation is the principal authority governing charges. Such authority is possessed by only a few agencies, such as NLM and the National Aeronautics and Space Administration. Other authorities governing charges are title V of the 1952 Independent Offices Appropriations Act (31 U.S.C. 483a) and OMB Circular A-25 (1959) (48).

Title V of the Independent Offices Appropriations Act requires that the Government set prices to recover as fully as possible the entire costs of providing a service, taking into account the public good and the benefit to the user. Indirect costs that benefit the public at large, rather than the purchase of specific services provided, cannot be included in the authorized fee.

OMB Circular A-25 requires that a reasonable charge should be made to each identifiable recipient of a Government service from which the recipient derives a special benefit. The identifiable recipient of a Government service who derives a special benefit above and beyond what the public at large received from the service should be charged, so that the Government recovers the full cost of rendering the service. Exceptions are allowed when the recipient is engaging in a nonprofit activity designed for the public safety, health, or welfare and when payment of full costs would not be in the best interest of the program. The circular prohibits Government agencies from charging more than the cost recovery level, since it limits the charge to the services' total cost and not the "value" of the services to the recipient.

Although both title V of the Independent Offices Appropriations Act and OMB Circular A-25 are concerned with setting prices for Government services in general—and not specifically information services—more and more, the policies are being applied to information products and services. Owing to their lack of specificity, interpretation of these two authorities is very difficult, and despite two Supreme Court cases, the questions of who and how much should be charged remain unclear. Indeed, in 1979 the General Accounting Office (GAO) recommended "that OMB should work with the executive departments to clarify the circular to state clearly when charges should be made and the manner in which full costs should be recovered" (55).

Operational considerations are also involved in setting Government prices. In general, an agency operating under a fixed budget will not attempt to recover costs in excess of its budget regardless of demand. A disincentive to increased productivity occurs because production costs and billing and accounting costs come from appropriated funds, but in most cases any funds collected by an agency are returned to the U.S. Treasury. The type of information good and service also influences the pricing practices of a federally funded information organization, and the organization usually has different criteria in establishing a price for a primary (e.g., a technical report), a secondary (e.g., a computer search of bibliographic information), or a tertiary information product or service (71).

As a result of the vagueness and variety of the laws and regulations governing pricing, a Federal information organization has many pricing options available. Government organizations are not established to raise revenues and therefore choose their pricing strategy on bases other than economic considerations.

The mission of the organization is a major factor in a Federal information organization's determination of the amount of costs and the categories of costs they want to recover. King (71) concludes that the Government has three goals in providing scientific and technical information: 1) assuring the distribution of the results of federally funded research or federally collected information; 2) supplying an agency and its grantees and contractors with the information they require for their mission; and 3) providing a particular community (i.e., health or education) with information, regardless of the origin of the information or the recipient's funding source.

Social utility of the information provided is another major factor in pricing. When the Government assumes the responsibility of allocating resources for the development and distribution of information, its pricing philosophy is premised on the societal benefit that information development and distribution bestows. Distributing information free would at first glance appear to realize the greatest social benefit, but there are costs to the Government associated with the development and distribution of information which reduce the net value to society. Thus, calculating a price which stimulates the greatest use of an information good or service, but discourages frivolous use, is another purpose of pricing (71).

The cost categories and methods used to calculate costs vary widely among federally funded information organizations, but in no case is there an attempt to recover the costs of creating information. The most frequent method of costing bibliographic computer searches, computer data searches, bibliographic data bases, and bibliographies is using fixed costs and costs associated with reproduction and distribution (71).

Considering the variety of methods used and the different aims of pricing, it is not surprising to find inconsistent pricing practices among Government information organizations. In 1979, GAO found that information organizations that provided scientific and technical information in five Government agencies in general were not functioning in accord with title V of the Independent Offices Appropriations Act and OMB Circular A-25 in their pricing practices.

In its analysis, GAO did not distinguish between those programs that have enabling legislation regarding pricing and those that do not, nor between data bases and on-line service. It reported that, for the most part, the agencies providing scientific and technical bibliographic services did not charge for providing such services, were inconsistent in applying cost recovery policies when charges were made, recovered less than 15 percent of the costs associated with user services, and did not equitably recover the costs of bibliographic data services suppplied to private organizations for commercial purposes. GAO attributes the inconsistent practices partly to the ambiguity of OMB Circular A-25.

Many of GAO's findings have been confirmed in other studies. A survey, released in 1981, of 24 human services clearinghouses that are partially or totally supported by Government funds found that 9 recovered no costs through user fees, and 14 recovered some (12). An exception was the National Technical Information Service (NTIS), which is required by law to be self-sustaining.

A 1980 analysis of 111 federally funded information organizations found that 43 percent did not charge their users for information services and products (71). Further, charging practices varied with the agency, the size of the organization's budget, the type of operating organization, and the type of service or product. The product and service most often charged for are books (42 percent), nonprint media (44 percent), computer bibliographic searches (39 percent), and computer data searches (41 percent). Thirty percent of the information agencies funded by the Federal Government charged user fees for their bibliographic data bases.

A Changing Emphasis in Federal Pricing Policy for Information Goods and Services: "Full Cost Recovery"

At this time, the pricing of data base tapes and bibliographic computer searches by a Federal information organization falls under the authority of the organization's enabling legislation, title V of the Independent Offices Appropriations Act and OMB Circular A-25. As just noted, the OMB Circular A-25 calls for instituting charges that should recover the "full costs" of rendering a service, but the vagueness of the language has left it open to considerable interpretation. In the past 2 years, OMB has emphasized full cost recovery in further directives, again without defining the term.

In June 1980, an OMB draft circular, "Improved Management and Dissemination of Federal Information," was issued for comment. One of its requirements was that information made available through other than the depository library information system be provided at a price that would recover all costs to the Government associated with disseminating that information—including printing, processing, and retention—but excluding the costs associated with the production or creation of the information. The circular was not issued in final form, because OMB felt that the new administration coming into office in January 1981 should set policy, and because some of its requirements are incorporated in the Paperwork Reduction Act (Public Law 96-511) of December 1980.

In the Reagan administration, the issue of recovering costs of information products and services has become an increasingly important one (7). In April 1981, OMB Bulletin 81-16 imposed a moratorium on the production and dissemination of certain audiovisuals and publications and called for recovering the costs of production through use fees. In fall 1980, the Government halted the production of many agency publications.

Of more direct interest is OMB Memorandum 81-14 released in September 1981. In implementing the 1980 Paperwork Reduction Act, OMB has requested that all Federal information centers be evaluated. Memorandum 81-14 sets forth criteria to be used in the evaluation by departments and agencies. One criterion is whether or not the information organization prices its products and services in order to recover their full cost. Officials at OMB recognize that "full costs" are not defined in the memorandum, but suggest that if full cost recovery for information goods and services were required, the definition would be flexible and would vary according to the specific case (7).

Also of interest is an unsuccessfully offered amendment to the 1981 Senate bill, S. 800, which reauthorized the Medical Library Assistance Act. The amendment would have required NLM to recover the full costs of products and services sold to domestic profitmaking institutions and foreign private users. NLM products and services sold to a nonprofit institution were exempt from full cost recovery only as long as the institution did not provide information services to profitmaking institutions. Full costs were defined as "the direct and indirect costs (including overhead) applying cost accounting principles associated with: (i) the administrative and intellectual preparation of information products; (ii) the creation and maintenance of systems for the storage, retrieval, and dissemination of these products: (iii) the storage and retrieval of these products; and (iv) the dissemination of these products in whatever form." Neither the 1982 Senate bill (S. 2311) nor the 1982 House bill (H.R. 6247) which propose reauthorization of the Medical Library Assistance Act consider full cost recovery.

The increased emphasis on full cost recovery results from a number of causes, including reductions in Federal expenditures and expectations that users of all Government products and services pay for the cost of the product or service. At the same time, members of the private information sector have become more insistent that the Government recover the costs of its information goods and services. The Information Industry Association, which speaks for many private sector firms in the field, has said that "provision of subsidized information services by Government at low prices (or no cost at all) is blocking and delaying the ability of the market economy in information to deliver low-priced information to everyone" (66).

On the other hand, some experts contend that the imposition of a full cost recovery policy on information products and services will limit an individual's choice by limiting financial access. Many, particularly the library community, feel that some social benefits of information will be lost under this policy and that full cost recovery would have "long lasting and deleterious effects upon equal access to Federal information for both the private and public sectors" (8).

The implications of instituting a full cost recovery policy for Government goods and services are not fully understood. Although full cost recovery appears to be a simple and straightforward calculation, the principle is nebulous and "must rely ultimately on arbitrary and economically indefensible accounting conventions" (21). The term full costs has not been defined with any degree of precision, and there are various notions of the cost categories to be included in calculating the full costs of information goods and services and the allocation of such costs (80). The problem is particularly difficult with respect to joint products, i.e., when more than one product is produced with the same resources.

Some observers believe that the expected increased revenues to the Government resulting from a full cost recovery policy might not materialize. The increased price that would be charged under a full cost requirement might result in a fall in quantity of goods and services sold and might lead to a loss of economies of scale, and the "alleged full cost price would fail to produce revenues equal to costs as it is intended to do" (21). A lower price may have benefits to the supplier—the Government—in the form of increased revenues, and benefits to the user in the form of lower prices.

A recent notion of including creation costs (i.e., the costs of creating information products and services) in a full cost recovery calculation may increase the price of Government-sponsored information products and services sharply. Perhaps even more than other cost categories associated with the production and the distribution of information goods, the costs associated with creating information goods are ill-defined. But the issue is acute because of pricing practices in the private information sector. Until now, private firms for the most part have not included creation costs in the leasing fees for their data base tapes but have absorbed the costs by overpricing the print products associated with the data base. It is questionable how long this practice can persist with changes in user patterns from print products to computerized information. If private firms include creation costs in costing their data bases, they may be put at further price disadvantage with respect to Government-sponsored computerized information products and services.

The copyright law also affects the principle of "full cost" recovery, primarily as it applies to the sale of print material and the leasing of tapes of the data bases, but not to the provision of on-line access to data bases.

The 1976 Copyrights Act (Public Law 94-533) provides no protection for any product—be it a report, data base tape, or geologic map—produced by Government employees on Government time. Such works are considered to be in the public domain; any domestic individual can copy them and sell them to anyone else.

A full cost recovery policy assumes OMB can establish a formula based on accounting principles to determine an appropriate price for Government documents, magnetic tapes, etc. *Simply* stated, such a formula would divide an agency's budget by the number of products sold in the year prior to instituting full cost recovery, yielding a per product price. The price would equal the average cost. The example below is purely hypothetical, because it assumes that the agency is producing only one product and not performing any other functions. It therefore does not account for the problem of allocating costs between products and among products and services.

If an agency's budget is $22,000 and it produces one report that is sold to 22 customers in year 1, in year 2, following the adoption of a full cost recovery policy, the agency will charge each customer $1,000 for the report. Because Government products are not protected by copyright laws against domestic copying, a single entrepreneur could buy one copy of the agency's report for $1,000, duplicate it, and sell it to the other 21 interested customers, or any interested customers, for less than the Government's price. The private entrepreneur's price could be lower because it does not include the high costs of creating and developing the product. As a result of the lower price of the private sector, the agency might lose its market for the product. Thus, in order to ensure full cost recovery, an agency would have to charge its first customer the full cost of production, in this case $22,000.

Theoretically, such a scenario is possible now, without an established full cost recovery policy. But because of subsidies, the Government price is less than the price a business could charge if it chose to reproduce Government products. Full cost recovery policy may effectively drive the Government out of the distribution business.

The effects of a full cost recovery policy could be mitigated if there were a clause in an agency's licensing or purchase agreement with a private firm that prohibited resale of its print or computer tapes product. It is not clear, however, whether such a restriction is enforceable. Since there is no copyright on Government products, the Government may not have the power in legal terms to hold buyers of its products to such a contract. Such a clause currently exists in NLM's licensing agreement for MEDLARS data tapes, which are priced much higher than others in the Government. It has not been challenged to date.

Thus, it appears that full cost recovery and other pricing principles for Government information products and services need a fuller and more comprehensive examination if all their ramifications are to be identified. There may be no one pricing formula that represents the correct price for all Government products and services, or for that matter for any one product or service.

NLM's Pricing Policies*

Introduction.—In almost all respects, NLM's pricing of its products and services is consistent with the National Library of Medicine Act, title V of the Independent Offices Appropriations Act, and OMB Circular A-25, congressional opinion, as expressed in hearings and reports, and the opinion of NLM's Board of Regents.

The National Library of Medicine Act authorizes the Secretary of Health and Human Services with the advice of NLM's Board of Regents to make publications, facilities, or services available: 1) without charge as a public service; or 2) on a loan, exchange, or charge basis; or 3) in appropriate circumstances, under contract arrangements made with a public or any nonprofit agency, organization, or institution. The purpose of this authorization is to advance the legislative mandate of the Library of making scientific and other information readily available in order to promote the Nation's health. Because the act offers the Secretary of Health and Human Services alternatives in setting charges, the Library has been able to be responsive to changing needs and has modified its pricing policies and practices from time to time.

---

*Specific prices for MEDLARS products and services are reported in ch. 2.

OMB Circular A-25 exempts Federal agencies from recovering the full costs of services when the recipient is engaging in a nonprofit activity designed for the public safety, health or welfare, or when payment of the full fee by a State, local, or nonprofit group would not be in the interest of the program. According to NLM officials, only 3 percent of MEDLARS users do not meet one or both of these conditions (56).

NLM has modified in its pricing policies over the years. The Library's initial free Carnegie library tradition of not charging for any of its products or services was changed with the advent and extraordinary growth of MEDLARS, which created a new and relatively expensive category of library activities. When the Board of Regents decided it was necessary to recover some of the Library's costs, it determined that the taxpayer should be responsible for basic library products and services but that other costs, particularly those associated with the new computerized products and services, should be borne, at least in part, by the user. The Library recognized that its appropriations could not support the entire costs of NLM's planned communications network, and so had its on-line system designed keeping in mind that it would recover a portion of the costs associated with providing its services from the users of the services (33).

The Library's policy in charging users has been relatively consistent. It is based on four assumptions that have varied only slightly from time to time. They are: 1) the biomedical community of users should share the cost of on-line services with NLM; 2) NLM should support the generation costs of building the data base and the users should pay the costs of accessing the system; 3) all users should have equal access to NLM services and all sectors of the user community should be charged the same amount for NLM's products and services; and 4) charges are imposed to provide a degree of management control over the rate of the system's growth and to make the service as independent as possible of NLM's appropriations (101).

Data Bases.—The NLM Board of Regents reversed a 1965 policy decision to withhold the sale of MEDLARS tapes for profitmaking purposes in 1970 on the basis that MEDLARS was no longer experimental and that OMB Circular A-25 required agencies to recover all or part of their costs by sales in developing products of commercial value (45). A fixed fee was established based on the dollar value of services performed by foreign centers in exchange for access to MEDLINE and other considerations.

In January 1982, the pricing structure was changed. The Board of Regents at their October 1981 meeting resolved to "endorse a change from a fixed fee for the MEDLARS tapes to a use fee rate structure" (102). The use fee has the advantage of providing information on data base use as well as serving a revenue purpose. The Board also recommended that it "continue to delegate to the NLM Director authority to adjust price structures in response to changing situations" (102).

Most other Government data base tapes are also leased on a use fee ratio structure. Characteristically, Federal agencies turn their data base tapes over to NTIS without charge. NTIS is then responsible for marketing the data bases, distributing tapes to customers, and establishing a price for their provision. The price of each data base is established by calculating its direct costs to NTIS, including management and marketing, and NTIS' fixed costs. If a new product or service is to be launched, NTIS considers the anticipated market value in its calculations. (See app. G for a discussion of AGRICOLA and ERIC, two data bases distributed by NTIS.)

On-Line Access Charges.—On-line access to MEDLINE was provided without charge from 1971 to 1973. User charges were adopted as a means of "ensuring that available NLM resources could continue to provide equal access to MEDLINE services, enabling the Library to sustain the quality and the performance of the system by an appropriate degree of control of the system and making it possible for the continued increase in numbers of outside users to be largely independent of NLM appropriations" (99).

In 1975, on-line access charges for all MEDLARS data bases, except TOXLINE and CHEMLINE, were raised. Later that year, a differential connect hour rate between prime and nonprime hours for all MEDLARS data bases was imposed in order to strengthen management controls over the use of the Library's computer system by more evenly distributing the workload, which was overburdening during peak hours. It was projected that usage would shift from prime time to nonprime time if the rate were lower during that period. The rates were raised as a means of attaining full cost recovery of those costs associated with NLM's provision of on-line services "outside the walls of the Library," such as telecommunications costs and backup computer costs. In 1980, connect hour rates to CHEMLINE, TOXLINE, and TOXBACK were raised to reflect increased royalty charges.

In October 1, 1981, on the advice of the NLM Board of Regents, connect hour rates charges for all domestic users were raised again, and comparable rates for foreign centers were established as of January 1982. The price per page printed is the same as before, because NLM calculated that the Library is recovering all costs for this process at this time (33). But the cost calculations for the new charges are very different from those used before. The Library redefined accessing costs to mean everything associated with on-line access to the data bases. Thus, an additional $1 million was added to the costs attributed to on-line services by including such items as overhead costs, computer costs related to the service, and the costs of managing the system. These latter costs were not included previously because they are appropriated costs.

Congressional Actions.—The House and Senate Appropriations Committees regularly review NLM's pricing policies and charges before they are formally adopted, and have agreed with them for the most part in the past. In 1974, Representative Flood expressed interest in putting MEDLARS on a more self-sustaining basis. Both committees have also expressed concern about the effects of cost sharing on the dissemination of health information, particularly to small institutions, and repeatedly have questioned NLM's Director on this issue (150,151).

Because the enabling legislation for NLM does not need periodic renewal, the responsible authorizing committees have historically not been involved with the Library's pricing practices. As previously noted, however, in 1981 the Senate Committee on Labor and Human Resources considered an amendment to the Medical Library Assistance Act (S. 800) that would have required the Library to recover the full costs of its products and services from profitmaking institutions.

Although the amendment was defeated, NLM is attempting to respond to the requests for full cost recovery expressed in it, by OMB, and by members of the commercial information sector. The Director of NLM established a task force as early as April 1981 to develop a system which would capture and allocate on a regularly prescribed basis all costs associated with NLM information services.

In the fall of 1981, NLM increased charges for on-line access to MEDLARS data bases and designed a new pricing strategy for the MEDLINE data base tapes. The Library maintains that it made both modifications to accommodate the public and private demands for change while maintaining its policy positions. At its October 1981 meeting, the Board of Regents reaffirmed its position against a differential pricing structure and for "equal access and equal charges for all users" (101). (The issue of differential pricing is discussed in ch. 6.) It also said "the cost of building data bases and housing them should be the Governmental responsibility of NLM, but accessing the system should be paid by the users" at full cost (101).

Chapter 6

MEDLARS and Private Health Information Systems: Discussion of Domestic Policy Issues

Chapter 6

MEDLARS and Private Health Information Systems: Discussion of Domestic Policy Issues

Advances in the application of computer and communications technologies to information transfer have focused attention on on-line systems in the public dialog concerning the Government's role in the creation and dissemination of information goods and services. Recently, the National Library or Medicine (NLM or the Library) has figured prominently in the debate. One view is that the achievements of its computerized bibliographic system, MEDLARS, indicate that NLM is fulfilling its mandate to create and disseminate health-related information. Another view is that MEDLARS' success prevents the growth of private sector organizations that create health-related data bases and commercial firms that vend on-line access to health-related data bases.

This chapter examines two sets of issues. One set concerns the range of NLM's computerized products and services. The Library regards MEDLARS as an extension of its library functions, while others say MEDLARS does not fall within traditional library functions and is an inappropriate NLM activity.

The second set of issues pertains to the pricing of NLM's computerized bibliographic products and services, specifically, NLM's fees for leasing NLM data base tapes and NLM's charges for on-line access to its data bases. NLM's position is that its pricing policies aid in the dissemination of health information to all who seek it; others argue that MEDLARS' subsidized prices give NLM a competitive edge over private sector firms and that NLM should recover the "full costs" of its computerized products and services. The debate concerning full costs remains active, although it appears that NLM is moving toward recovering the full costs of its computerized products and services. Nevertheless, as will be discussed, the definition of full costs is open to various interpretations.

The issues examined in this chapter are considered within a general framework of the Government's role in the allocation of resources to information development and distribution; the effect of the Government's involvement in information activities on certain segments of the private information sector, and the health community; and the historic role of the Government in health information activities (see ch. 5). Specific criteria used to examine the issues are product differentiation, historic precedence, the presence of positive externalities (the social benefits received from a product or service exceed the sum of the benefits received by separate individuals), Governmental costs, present or potential private sector involvement, the effects of private sector participation on the creation and dissemination of computerized bibliographic health-related information, and international implications. (International issues are discussed in app. I.)

Although the analysis in this chapter focuses on current issues, that focus is not at all intended to minimize the importance of new and emerging technologies on biomedical information policy. Current issues are likely to be altered by changes in computing and communications technologies. Thus, the analysis in this chapter should be considered in the context of information concerning future information technologies. (These technologies are discussed in app. H.)

RANGE OF NLM'S COMPUTERIZED PRODUCTS AND SERVICES

Data Bases

MEDLARS, NLM's computerized bibliographic retrieval and technical processing system, includes MEDLINE (NLM's original data base) and many other health-related data bases. The rationale for NLM's creating or maintaining data bases varies from data base to data base, because of their diversity. (See ch. 2 for a detailed description of MEDLINE and, NLM's other health-related data bases.) MEDLINE is the focus of this chapter's discussion. For the most part, the other health-related data bases are used to illustrate specific points: it is not possible to generalize about them and it is beyond the scope of this report to assess individual ones.

The immediate concern with NLM's creating and making available health-related data bases arises from the claim that NLM is competing with the private sector's creation of such bases. Basic to the concern is the relation of NLM's health-related data bases to those created by the private sector. Do these data bases provide identical, similar, or complementary information? Can one base be substituted for another?

It is generally accepted that MEDLINE and "similar" health-related data bases do not duplicate one another, but that there is a degree of overlap. Overlap in coverage can occur at two levels. Journal overlap occurs when two or more data bases contain articles from the same journal. Journal article overlap occurs when two or more data bases contain citations of the same article from the same journal. Journal overlap does not necessarily indicate journal article overlap (112), because different data bases may include citations of different articles from the same journal. Moreover, the contents of the same article are often analyzed differently for different data bases in order to meet user needs. For example, an article on the biologic differences between two stages of a particular cancer can lend itself to a variety of analytic approaches centered on basic biologic processes, or on diagnostic techniques, or on chemical analysis, or on clinical manifestations of the disease.

EXCERPTA MEDICA and BIOSIS PREVIEWS are the data bases that appear to be the most similar to MEDLINE in subject coverage, but the extent of overlap among them has not been accurately determined. An early study comparing the results of searches on biological, medical, and veterinary subjects in EXCERPTA MEDICA and MEDLARS data bases reports that of the 226 relevant references retrieved from EXCERPTA MEDICA data bases and 467 relevant references retrieved from MEDLARS data bases, only 94 references (or 13.6 percent) were found in both secondary sources. The authors' tentative conclusion is that the systems complement each other, since if one system answers a question poorly, the other system answers it well (154).

A similar conclusion is reached in a later study that compared the two on-line search files as they existed in 1978 (19). An estimated 42 percent of the on-line MEDLINE file was covered also by the EXCERPTA MEDICA file and about 31 percent of the EXCERPTA MEDICA file was covered by the MEDLINE file. The methodology employed an author search: a search using indexing terms might have disclosed a different degree of overlap. The author concludes that "given the relatively large numbers of unique records contributed by each file, they are clearly complementary services and any comprehensive search should make use of both files" (19).

MEDLINE and BIOSIS PREVIEWS were included in an analysis of 14 major scientific and technical data bases for both journal and journal article overlap.[1] Among all the 14 data bases, there was approximately 20 percent journal overlap, and among these journals there was 23 percent journal article overlap—i.e., only 4.6 percent of all the articles in all the journals cited in any one data base were cited in more than one data base. The authors concluded that the amount of overlap was much less extensive than they hypothesized before the study began. In any event, overlap is often useful, particularly when comprehensive retrieval is important, and MEDLINE and BIOSIS PREVIEWS have been used successfully in such a complementary fashion.

---

[1] T. C. Bearman and W. A. Kunberger, *A Study of Coverage Overlap Among Fourteen Major Science and Technology Abstracting and Indexing Services* (Philadelphia: National Federation of Abstracting and Indexing Services, February 1977).

The data bases differ in other ways, including the professional fields of interest and focus of each of their abstracting and indexing services. For example, MEDLINE includes journals in nursing, dentistry, and allied health fields; EXCERPTA MEDICA does not. On the other hand, EXCERPTA MEDICA includes materials related to other health-related disciplines not found in MEDLINE, as do BIOSIS and the other health-related bases. The bases also differ in their coverage of English language publications. Sixty-five percent of the journals indexed for MEDLINE are in English, 60 percent of the journals indexed for BIOSIS are in English, and 50 percent of the journals indexed for EXCERPTA MEDICA are in English. Other differences include the percentage of abstracts in each data base.

There are few rigorous studies of overlap concerning other NLM health-related bases and "similar" data bases created by the private sector. One analysis of toxicology information abstracting and indexing services found that each service contributes somewhat differently in terms of scope of coverage and type of coverage, and that no one service is comprehensive or exhaustive in the field (112). The study points out the advantages of diversity of biomedical data bases and printed products, particularly in the toxicology and chemical fields.

Product differentiation, however, does not completely negate the possibility of substitutability among health-related data bases or printed indexes and abstracts, and at this time, there are no hard data on the extent to which the existence of one influences the use of another. Some claim that libraries with limited budgets, such as hospital libraries, choose MEDLINE instead of "similar" data bases on the basis of price alone. Others claim that price is only one criteria and that hospital libraries choose MEDLINE instead of similar data bases because the literature cited in MEDLINE is more relevant to their needs: the practitioner who uses the hospital library will choose MEDLINE over other data bases because its content and scope are more related to clinical medicine (155).

The little evidence currently available suggests that the citations in MEDLINE, and other NLM health-related data bases, do not duplicate those available in any one base currently created by the private sector, and that the availability of a diversity of biomedical data bases may be advantageous. It is on this basis that the following arguments for and against NLM's creating and making available MEDLINE, and other health-related bases, are presented.

NLM created MEDLINE as a result of the computerized production of *Index Medicus,* and the medical and information communities throughout the world perceive this data base as an extension of the printed publication.* The printed index has a long and respected history, and is used worldwide: of the 5,888 copies sold in 1981, 2,623 were sold abroad. MEDLINE is also used extensively overseas. Foreign centers obtained access to the data base just shortly after it became available in the mid-1960's through quid pro quo arrangements (see app. I). In this country and abroad, the medical and scientific communities depend heavily on *Index Medicus* and MEDLINE and, with few exceptions, support their continued production.

The historic relationship between *Index Medicus* and MEDLINE in conjuction with NLM's legislative mandate to publish and make available "catalogs, indexes and bibliographies . . . to the material it collects" (Public Law 84-941) is an important argument in favor of NLM's continuing to create and make available MEDLINE. The argument is strengthened by the joint production of the two products. *Index Medicus* and MEDLINE are tied together by the production process, as the computer tape which is accessed in MEDLINE serves as a vehicle for the production of *Index Medicus* and derivative products. Indeed, the computerization process that results in MEDLINE is the least expensive way of producing *Index Medicus.*

---

*There is little, if any, expressed interest in NLM's discontinuing the creation of *Index Medicus.*

The presence of social benefits (externalities—see ch. 5) is another criterion by which to assess NLM's continued creation of MEDLINE and other health-related data bases. Insofar as it is believed that the general public health requires expenditures for information products and services which individuals will not incur for their own benefit, Government provision of such information may be necessary. More specifically, if MEDLINE and NLM's other bases provide information to physicians, researchers, and other health professionals that results in an improvement in the health of various members of the general public, society may want the information created and disseminated, even if the physicians, researchers, and other health professionals do not perceive that they get enough benefit from the information to pay the price of buying it on the private market.

In considering how far such conditions exist for MEDLINE and NLM's other health-related data bases, it is assumed that some of the primary literature cited in the data bases contains information that results in an improvement in health. But do secondary sources, which refer to the primary literature, contain information that is needed to protect the general public's health? They would appear to, since the primary literature is voluminous, scattered in increasing numbers of journals, and virtually inaccessible without a well-constructed index (see app. D). Thus, secondary information is considered to have social value because it "leads to the use of primary information and can reduce costs of identifying and locating information for the Government and other users" (71).

The more significant questions are whether MEDLINE and NLM's other data bases cite the primary literature that contains information to improve health, and whether the citations are readily available to institutions and individuals that benefit the public's health. There is no simple answer to the questions. It is likely that the public receives the most immediate benefit indirectly from information that is available to practicing physicians and other health practitioners. MEDLINE contains references to literature in many fields that contribute to the science and practice of medicine and public health, with an emphasis on the basic medical sciences and clinical medicine. While there is evidence that MEDLINE is responsive to the needs of practitioners (24), there are practitioners who suggest its responsiveness could be improved.

Because of subject content, subject headings used in analyzing articles, and the organization of the headings, searchers do not always easily retrieve references to the literature that physicians practicing in patient care disciplines and settings find useful. This is not to say that *Index Medicus*/MEDLINE is not used by practicing physicians and other health practitioners providing patient care. Indeed, it is the most used secondary source of all bibliographic biomedical sources (135), and there is extensive anecdotal and indirect evidence of its use by patient care practitioners.

On the other hand, information required by biomedical researchers is identified and organized in MEDLINE for easy retrieval. If one believes that biomedical research benefits the general public health, then one may very well believe that NLM is justified in producing MEDLINE. Indeed, Congress established NLM to "aid the dissemination of scientific information important to the *progress* of medicine and to the public health" (emphasis added) (Public Law 84-941).

NLM's other health-related bases require individual evaluation with respect to the presence of social benefits. Each contains citations of literature that varies in form and content: a few of the data bases contain numeric or representational data. The category of user varies from data base to data base as well. For instance, TOXLINE contains references to information that has an immediate, or delayed, beneficial effect on the public health. But 50 percent of its usage, as measured by connect hours, is by commercial firms, including drug companies. It could be argued that commercial enterprises receive enough personal benefits from the information they receive to pay the cost of buying it on the market.

The same reasoning can be applied to CHEMLINE, which receives 50 percent of its usage from commercial firms, but not to HEALTH, POPLINE, BIOETHICSLINE, and other bases. For example, only 1 percent of the approximately 318 hours of HEALTH usage in 1981 was by commercial firms, while 80 percent was by hospitals, allied health and medical schools, and research institutions. Whether such users receive sufficient personal benefits from the information to pay for it on the market at current levels of usage is in doubt.

Related to the above discussion of social benefits is the Government's allocation of resources to health-related research, since the Government funds research because of the social benefits derived from its findings. The Government supports more than 50 percent of the biomedical research conducted in the United States, and over the years has noted the importance of organizing and disseminating the results of biomedical experimentation. (See the discussion of the history of health information in ch. 5.) It sometimes pays charges to professional journals to publish research results, and it has developed and maintains an organizing tool, *Index Medicus*/MEDLINE, for accessing the literature on Government-sponsored and other research. If the Library were to stop producing these bibliographic sources, it could be argued that meeting the goals and objectives of the Government's biomedical research policy would be made more difficult.

On the premise stated above, as Government decreases its funding for a type of research, its justification in sponsoring the results of that research is weakened. To illustrate, HEALTH is a data base that centers on the health services research and health care delivery literature. The appropriations for two of the major Federal organizations that sponsor health services research, the Office of Research and Demonstrations and the National Center for Health Services Research in the Department of Health and Human Services, decreased from $69 million in 1980 to $40 million in 1982. Should the Government, therefore, decrease or discontinue its support of HEALTH?

The issue is extremely complicated and must be measured against other criteria. The data base is used by hospital administrators and health practitioners as well as health services researchers. As Government expenditures for health care continue to increase annually at enormous rates, and to the extent that health service research suggests ways to contain health care costs, the rationale for Government sponsorship of access to this information is strengthened.

The private sector might be inclined to create MEDLINE and NLM's other health-related data bases with the same degree of consistency, quality, and comprehensiveness as NLM has. But the desire or lack of desire of the private sector to do so is purely speculative. At the present time, no established private concern has expressed an inclination to undertake the creation and maintenance of MEDLINE. It would be difficult for a private concern to rationalize the expenditure of money required to create and maintain the data base, since it requires the acquisition of a large collection, library staff to index and process the collection, and other capital-intensive equipment and facilities. Nonetheless, venture capital is available for many information activities. The possibility exists that if NLM were to announce now that it would cease the publication of *Index Medicus*/MEDLINE in 5 years or so, some organizations might be interested in the project (137).

However, there might be potential disadvantages associated with private sector creation of MEDLINE, including the potential risk of not being able to follow through on the proposed substitute system. There would be no assurances of continuity, quality, and comprehensiveness. It is doubtful that another organization could establish a relationship similar to the one the Library has with the medical, research, and library communities. The interaction is of great value in developing and maintaining the quality of *Index Medicus*/MEDLINE in areas such as developing medical subject headings (see app. E).

If *Index Medicus*/MEDLINE were to be produced by the for-profit sector, the data base might not be as comprehensive. MEDLINE contains rarely used citations to articles in journals with low circulation and citations of possible future use. For-profit firms tend to be more selective and utilitarian in their operations, and might concentrate on common information and information of immediate use. The uncertainty of future demand may limit the information to be preserved by the for-profit sector, although the potential societal value of the rarely used information makes its preservation important.

Finally, the price to the user of *Index Medicus*/ MEDLINE might be higher if it were prepared by a private firm. The implications of higher prices are discussed below in the section entitled "Pricing Issues."

It is not possible to generalize when speculating about the private sector's inclination to create and make available NLM's other health-related data bases, because of their number and the diversity of their contents. However, the number of records in NLM's other bases and usage are low compared with MEDLINE (see ch. 2) and they may or may not be profitable for a private firm to create.

In September 1981, users interacted with MEDLINE for more than 5,500 hours, but they used TOXLINE, the next most utilized data base, for only 826 hours during the same period and CHEMLINE for only 285 hours. Nonetheless, in fiscal year 1981, with the NLM user charges then in effect, NLM recovered 110 percent of its accessing and tape costs associated with providing online access to TOXLINE and 75 percent of its accessing and tape costs associated with providing on-line access to CHEMLINE. If NLM's 1982 user charges had been in effect, NLM would have recovered 126 percent of its accessing and tape costs associated with providing on-line access to TOXLINE and 126 percent of its accessing and tape costs associated with providing such access to CHEMLINE (56).

In September 1981, BIOETHICSLINE and CANCERPROJ received only 31 hours of usage each and would not recover costs with such usage. But some data bases may reach a wider audience if produced by the private sector. For-profit enterprises generally have much more sophisticated marketing technique than the Government. Data base size and current usage are only two of many factors to consider in determining if a data base would bring in revenues. A few others include the availability of source material, the costs of production, and the possible publication of a print product from the data tapes. The last is essential, because data bases are not profitable currently unless produced along with a hard copy publication. Each of the factors varies with the individual MEDLARS data base.

If the commercial sector were to find it profitable to produce some of NLM's data bases, NLM might be responsible for creating and distributing only those bases that do not meet the market test. In that case, there is a chance that a perception might develop that NLM is creating and distributing data bases of little value.

On-Line Services

In addition to creating MEDLINE and other bases, NLM provides on-line access to them. Although the service is relatively new with respect to other NLM operations, it is a very visible activity and is associated with NLM's history and leadership role in bibliographic retrieval systems both here and abroad. It was established before commercial services provided on-line access to NLM's or to most other health-related data bases.

Currently, two commercial information services (vendors)—DIALOG Information Services, Inc. (DIALOG) and Bibliographic Retrieval Services (BRS)—lease the tapes of MEDLINE and a few other NLM health-related bases and vend online access to them. Are NLM, DIALOG, and BRS providing services in the same market? If NLM's on-line services are sufficiently different from those offered by the private sector, they may be operating in different markets. It is generally recognized that the private sector search services, as a result of better services and advanced software, are more efficient than those currently available from NLM (see ch. 4). It is not clear that NLM and private information services are different enough to conclude that the services are operating in different and discrete markets.

Another way of defining discrete markets is by the types of users being served by similar products or services. The markets are discrete if the users are sufficiently dissimilar. On this premise, there is some, but inconclusive, evidence that NLM and DIALOG, one of the private information services that sells access to NLM data bases, are providing services in different markets. As measured by connect hours, the majority of users that have online access to MEDLINE through NLM are in hospitals (40 percent) and academic institutions (19 percent). Although the exact figures are proprietary, there are indications that a large percentage of users that access MEDLINE through DIALOG are in commercial organizations. There is some overlap, since 10 percent of the users that access MEDLINE through NLM are commercial organizations, and 28 percent of DIALOG users are in academic and other nonprofit institutions. Nonetheless, DIALOG appears to be providing services to a market different from that of the Library. However, NLM and BRS appear to be serving the same market, as the majority of BRS users (55 percent) are in medical schools and academic institutions. But the boundaries of the markets are not necessarily firm and might shift if NLM were to stop providing on-line services.

Only very indirect evidence is available with which to assess the effect of NLM's provision of on-line services on commercial information services (vendors). Computer-based information services, a very much larger category of business firms than the three information services discussed in this report, realized almost $8 billion in revenues in 1979 and anticipate increasing their revenues by 29 percent in 1980 (42). Another indication is that BRS was recently acquired by a multinational firm, Thyssen-Bornemisza, for a sum higher than BRS' original capital investment. However, in May 1982, BRS said NLM's new pricing structures did not allow BRS to vend MEDLINE profitably (46). (This issue is discussed further in the next section.)

NLM and others are concerned that serious damage to the integrity of NLM's data bases might result if NLM were to discontinue providing on-line services. Because of on-line access, user training, and user services, NLM continuously interacts with MEDLARS users. The users provide information about any inadequacies of the data bases, thereby facilitating another quality check of the data bases by NLM. A counter argument is that NLM might find other ways to communicate with users if it did not provide on-line access to NLM's data bases. Among others, BIOSIS creates a quality data base without providing on-line services.

A parallel factor is the collection of quantitative data on users. The Library requires such data, if it is to create bases that reflect user needs. No matter who provides the search service—NLM or private sector firms--they will be precluded from obtaining comprehensive statistical data on users because of the privacy of the search process. With that limitation in mind, it would be possible for private information services to obtain and provide only general statistical information on users for NLM's use, for a fee.

A most important factor to weigh in assessing NLM's provision of on-line access to its data bases is the Government's cost of providing the service. The additional cost to NLM of providing on-line access is relatively low because of user charges. For fiscal year 1980, according to NLM, approximately $400,000 in appropriated funds would have been saved if NLM had not provided on-line access to MEDLINE but had continued to create the data base. Even if NLM were not to provide on-line access to MEDLINE, creation and maintenance of the data base would still be required. Service support, data base testing, disk storage, and computer hardware would still be maintained in order to provide the data base to vendors (103).

Somewhat higher costs related to NLM's provision of on-line access to its data bases are reported for fiscal year 1981 (56). In fiscal year 1981, NLM incurred costs directly assignable to offering MEDLINE publicly for on-line searching of $3.241 million. NLM recovered $2.336 million in user fees (including $42,000 in services by NLM for MEDLARS carrying out its basic library activities). Thus, it cost the Government $905,000 to provide on-line access to MEDLINE in fiscal year 1981, or 28 percent of the accessing and tape costs associated with providing on-line access. The accessing and tape costs associated with providing on-line access to all the NLM data bases were $4.757 million. Of this, NLM recovered $1.271 million, or 28 percent, in user charges. If NLM's current (higher) user charges had been in effect during 1981, the provision of on-line access would have cost the Government only $166,000 or 6 percent of the costs associated with providing access to MEDLINE and only $235,000 and 5 percent of the costs associated with providing access to all NLM data bases (56).

PRICING ISSUES

Pricing issues related to leasing NLM's data base and pricing issues related to NLM's on-line services receive separate consideration in the following pages. But these pricing issues are interdependent, because the leasing fee that private information services pay for NLM's data base tapes is a factor in information services' pricing of on-line access to the data bases. The relationship between leasing fees for NLM data base tapes and NLM's on-line access charges illustrates the heterogeneity and diverse interests of the private information sector. If NLM's leasing fees for its data base tapes are high, private sector firms that produce health-related data bases presumably will be in a better position to market their product. This will especially be true if the leasing fees for NLM's data base tapes are higher than the leasing fees for the tapes of privately created data bases. At the same time, however, private information services that lease the NLM tapes and subsequently vend them will have higher costs in providing on-line services.

The different effects on different members of the private sector are reflected in the current dialog concerning tape leasing fees for NLM's data bases. NLM's current leasing fees for MEDLINE tapes are new, having become operational as of January 1982. Thus, it cannot be determined whether the full costs of creating the data base are recovered by the leasing fee, although the fee is much higher than the costs of reproducing the data tapes and much higher than the leasing fees for almost all other Government-sponsored data base tapes.

Some data base producers contend that the NLM fee for leasing data base tapes should recover the full costs of creating the data base. On the other hand, information services (vendors) in the private sector feel that NLM's leasing fees are currently too high, and some propose that NLM follow the model used by the National Agricultural Library for AGRICOLA and by the National Institute of Education for ERIC (see app. G) (137). The tapes of ERIC are leased by the General Accounting Office (GAO) at the cost of reproduction: the tapes of AGRICOLA are leased by the National Technical Information Service (NTIS) at $1,220 for domestic use plus a fee. (See table 12 for NTIS leasing fees for Government data bases.) Both computerized data bases have a wide offering, from in-house systems of computers at universities, companies, and other organizations, and are offered by commercial vendors at relatively low rates (see table 8 in ch. 4).

Fees for Leasing Data Base Tapes

The principal pricing issues concerning the leasing of NLM's data base tapes are: 1) whether leasing fees are to recover the costs of reproducing the tapes only, or to recover both the costs of reproducing the tapes and the costs of creating the data bases; and 2) whether foreign lessees are to pay a different fee from domestic lessees.

One view is that leasing the data base tapes at the relatively low cost of reproducing the tapes would allow for private sector participation and at the same time widen the distribution of health information. Another view is that private producers of health-related data bases could be adversely affected by this practice. The American Psychological Association, for instance, claims its data base cannot effectively compete with a base prepared by the National Institute of Mental Health, which leases its data base tapes to commercial information services at the cost of duplicating the data tapes (9). Another argument is that all U.S. taxpayers pay for the data bases, but only a few commercial information services would benefit monetarily from leasing the data base tapes based on reproduction costs only.

The private sector, in general, does not lease its machine-readable data base tapes to include the costs of creating the data base. If creation costs were included, the on-line access cost would be so high as to discourage usage. For the most part, the print products associated with the bases subsidize their production. The use of machine-readable data bases, although extensive, is new; the user communities in the health fields are not quite prepared to pay high costs for information (69). In addition, data base production is in a period of great transition, with a declining demand for print materials and new technological develop-

Table 12.—Selected Data Bases Distributed by NTIS, December 1981

Department	Agency	Data base	Acronym	Annual lease	Use fee
Agriculture	National Agriculture Library	The Agriculture On-Line Data Base	AGRICOLA	$1,220 domestic 1,820 foreign	Yes
Commerce	National Oceanic and Atmospheric Agency	Aquatic Sciences and Fisheries Abstracts	ASFA	2,500	Yes
	U.S. Patent and Trademark Office	Patent Full Text File		10,920 domestic 16,380 foreign	Foreign only
		Patent Bibliographic File		5,460 domestic 8,190 foreign	Foreign only
		Patent Classification File		900 domestic 1,150 foreign	None
	National Technical Information Service	NTIS Bibliographic Data Base	NTIS	5,350 domestic 7,150 foreign	Yes
Defense	Defense Technical Information Center	Technical Abstracts	TAB		
Energy	Technical Information Center	Energy Research Abstracts	EDB	2,950 domestic Not available—foreign	Yes
Health and Human Services	National Library of Medicine	MEDLARS On-Line	MEDLINE	20,000	Yes
Interior	Office of Water Research and Technology	Selected Water Resources Abstracts	SWRA	2,300 4,600	Yes
	U.S. Fish and Wildlife Service	Pacific Island Ecosystem	PIE	500 750	Yes
NASA	Scientific and Technical Information Office	Scientific and Technical Aerospace Reports	STAR		
	Smithsonian Science Information Exchange	Notice of Research Project Data Base	SSIE	8,000 16,000	Yes

SOURCE: National Technical Information Service, 1982.

ments occurring rapidly. Accordingly, it appears unlikely that the private sector will lease its data tapes at a price that includes the costs of creating the data bases in the near future.

The problem is further complicated by the quid pro quo arrangements that NLM has with foreign centers. In exchange for either lease of the data base tapes from NLM or access to them by means of telecommunication to the NLM computer, these centers contributed almost $500,000 worth of services, mainly indexing services, in 1980 and more than $600,000 in 1981. If NLM were to provide the data base tapes at reproduction costs to the foreign centers, there would be a large drop in the quantity of indexing performed for the Library. Although no money is received by the Library in this transaction, existing or additional appropriations would be required by NLM in order to maintain the same level of indexing. Whether the foreign centers would agree to be responsible for the indexing of the foreign literature under a non quid pro quo arrangement is strictly conjectural.

NLM's quid pro quo arrangements with foreign centers have additional advantages. A rough examination of the MEDLINE citations in recent years shows that the percentage originating in the United States has remained relatively stable at about one-third of the total biomedical serial literature indexed (133). Furthermore, NLM's quid pro quo exchange agreements are quite important for the future, because the costs of publishing overseas have continued to remain below those of the United States and the probabilities of using electronic methods for dissemination of research results may reduce the U.S. percentage of MEDLINE citations. With a large percentage of medical journals being published overseas, the cooperative international system, mostly among governments or quasi-governmental organizations, benefits U.S. citizens as much as, if not more than, citizens of other countries.

A suggested alternative pricing policy to the current policy is to lease the data bases to private enterprises in the United States at the cost of reproducing the tapes, but to lease them to foreign countries to recover the full costs of their creation, i.e., differential pricing. On the surface this idea appears attractive. If the data bases are considered to represent a national investment, it seems appropriate that U.S. citizens benefit from the investment, but that foreign nationals not necessarily share in the benefits.

However, the use of differential pricing might have serious international policy ramifications. Foreign nations might view this change in pricing policy as an antagonistic act on the part of the United States. They might conceivably reconsider their arrangements with the Library, thereby interrupting the flow of some medical information from abroad. In addition, technical considerations reduce the possibility of enforcing differential pricing for domestic and foreign users of the on-line computerized system. It is very difficult to distinguish between domestic and foreign users, and foreign sources often find a domestic source of information where price barriers exist. Unauthorized copies of data base tapes are also feasible with new technologies.

On-Line Service Charges

If NLM continues to provide on-line access to its health-related data bases, several specific pricing issues need to be addressed are: 1) whether to establish charges for on-line access to the data bases to recover the full costs of accessing the data base; 2) whether to establish charges for on-line access to the data bases to recover the full costs of accessing the data bases and of creating the data bases; and 3) whether to subsidize MEDLARS or MEDLARS users. The issues are discussed here in the context of structuring charges for on-line access to MEDLARS data bases to best serve the public purpose for which NLM was created without inhibiting the operations and development of private information enterprises. The debate on these issues has been complicated even more by lack of definition of the term "full cost recovery."

There are advocates of guaranteeing all citizens free and equal access to publicly funded library and information services (161). Conversely, there are advocates of full cost recovery of Government information products and services, including the costs of creating the information. In establishing the prices for on-line access to MEDLARS, NLM is between these two positions. (See ch. 5 for the history of NLM's pricing policies.) It has adhered to the policy of sharing the costs of on-line services with the user community, with the taxpayers assuming the costs of creating the data bases and the user community assuming the costs of accessing the system.

Within the above framework, however, NLM has shifted its emphasis on the specific goals of its pricing structure. An economic analysis of NLM's on-line services in 1974 and 1975 investigated alternative pricing strategies and their relationship to social benefit (81). The study recommended marginal cost pricing for access in order to benefit the health of the Nation's people. It supported pricing access to MEDLINE at $8 to $15 per connect hour, which was the estimated price of recovering the full costs of that access, defined at that time by NLM as costs associated with on-line access "outside the walls of the library," such as telecommunications and backup computer costs.

In October 1981, NLM increased its price for on-line access to MEDLINE to $15 to $22 per connect hour. However, it did not justify the increase on the basis of social benefit as well as on the basis of full cost recovery of costs, as it did in 1975. Rather, the justification was on the basis of full recovery of access costs and on the basis of bringing on-line costs more in line with accessing other Government data bases and other discipline oriented data bases. The premise of the Library's cost calculations was also different, as the Library redefined access costs to include "*all* related computer and communication charges, *all* direct labor for system and network maintenance, *all* printing or duplicating charges, overhead and indirect costs, and training costs" (103) (emphasis added). NLM excluded from the cost base those costs which it would necessarily incur in maintaining the bibliographic apparatus of a national library, including capital costs.

According to NLM's definition of the "full cost" of accessing MEDLINE (i.e., the costs of accessing the data base but not the costs of creating the data base), and on the basis of NLM's analysis, NLM appears to be approaching full cost recovery for accessing MEDLINE on-line (103).* A recent analysis (56) confirms NLM's conclusion that it is arriving at full cost recovery for accessing MEDLARS data bases. For fiscal year 1981, NLM recovered 72 percent of the accessing and tape costs associated with providing on-line access to MEDLINE and to all the other NLM data bases. This analysis was based on NLM's charges for assessing MEDLINE which had been in effect before 1981. If NLM's current (higher) charges had been in effect during fiscal year 1981, NLM would have recovered about 94 percent of the accessing and tape costs associated with providing on-line access to MEDLINE and and about 95 percent of those associated with providing on-line access to all of NLM's data bases. NLM's method of identifying and allocating costs was in general considered reasonable for achieving order to achieve full cost recovery of its accessing costs (56).

It is too early to evaluate the effect of the increased charges for accessing MEDLINE on the private information sector. The lower pricing structure for on-line services that was in place until October 1981 did not appear demonstrably to have impaired the growth of the industry. It may indeed have fostered its development. As the first to offer computerized services in the medical bibliographic field, NLM with its low prices may have been instrumental in breaking down initial resistance to new products and services, thereby opening up markets for private firms (137,163).

In any case, there are problems with generalizing the effects of NLM's charges on the private information services (vendors). Not only are the firms diverse, with sometimes conflicting interests, but the relationship between leasing fees for NLM data base tapes and on-line access charges masks the effects of the latter. For example, BRS was able to successfully vend MEDLINE at the same price as NLM under NLM's pre-October 1981 prices for on-line services and NLM's pre-January leasing fee structure. As of January 1982, NLM changed its leasing fee structure from a flat annual fee to an annual lease fee plus a $4 per hour usage fee and a $0.01 per citation print fee all of which BRS passes through to users. NLM, however, does not impose a per hour usage fee or a per citation print fee on users who access MEDLINE on NLM computers. Because BRS customers are high-volume users and BRS provides on-line access to MEDLINE in approximately the same market as NLM, BRS finds it can no longer provide access to MEDLINE at the same price as NLM does, and obtain a reasonable profit. Thus, BRS is concerned that its users will switch from its service to NLM's and considers NLM's current leasing fee structure and on-line service charges unfair (46).

On the other hand, NLM's increased charges for on-line access to MEDLARS data bases are now more in line with DIALOG's charges. DIALOG has more of MEDLINE on-line than either NLM or BRS, and thus incurs higher costs which are reflected in its prices. DIALOG does not seem to be unduly affected by NLM's new leasing fee structure: the majority of DIALOG's users appear to be commercial organizations, and the majority of NLM's users are not (137).

Any assessment of the relation between NLM's increased on-line access charges to its data bases and the commercial viability of similar health-related data bases produced in the private sector would be based on very little evidence. The only information available are the statements of one or two private data base producers that, even at the current higher rates, NLM's charges for accessing MEDLARS in conjunction with NLM's leasing fees are detrimental to the sale of their products. Since financial data on private companies are proprietary, there is no way to substantiate the statements. As noted earlier, MEDLINE and similar bases are complementary and not identical, but there are no data on the extent to which the presence of one in a market influences the use of another.

---

*The rationale for NLM's cost allocation which leads to this conclusion is that each cost for fiscal year 1980 was allocated entirely or in part to one of the five categories: 1) essential national library services (ENLS), 2) data creation, 3) publications, 4) on-line services, or 5) tape distribution. ENLS were considered to be those activities, such as selection and acquisitions, technical processing, reference services, etc., that are basic and essential for a national library to perform its mission. These are services that would be performed by a national library whether or not there were products such as publications, on-line services or tape distribution. The cost of indexing journal articles was considered to be partly basic service and was divided equally between ENLS and data base creation. Costs assigned to data base creation were those which were incurred in gathering or preparing the data on which the publications, on-line services, or tape distribution are based but which cannot be specifically assigned to one of the three. Each cost which could be allocated specifically was appropriately assigned. Portions of salaries, benefits, and other costs associated with personnel (travel and training) were assigned according to an estimation of the time spent.

A new issue in calculating the full costs of accessing NLM's data bases is whether or not to include the costs of creating the data base as well as the costs of accessing the system in the calculation. The issue is of unique interest, mainly because it is of such recent origin. Until the unsuccessfully offered cost recovery amendment of the Health Care Research and Research Training Amendments of 1981 (S. 800), the creation of information was not referred to in any Government directive or practice regarding the pricing of information products or services: even the 1982 GAO report (56) on MEDLARS does not consider the costs of creating the data bases in its pricing analysis of NLM's on-line services.

Principle 6 of the report of the Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science (96) recommends that pricing policies for distributing information by the Federal Government should reflect the true cost of access and reproduction; however, the costs of data base creation are not included in the recommendation (see ch. 5 and app. F). The concept of creation costs is vague because the costs of creation can be considered indexing or abstracting costs only or can include acquisition costs and overhead costs as well. As mentioned above, private data base producers do not include creation costs in pricing the leasing fees of their data base tapes, and accordingly creation costs are not included in the on-line access charges set by commercial information services.

Most important is that a mandated function of NLM is to acquire, process, and index health information—and funds are appropriated for that purpose. It is questionable whether MEDLARS users should pay for a process that is a general library function which benefits all users of the Library.

The major advantages of including data base creation costs in NLM's charges for on-line access to MEDLARS data bases is the increased revenues to the Government and the possible beneficial effects to the private sector producers of similar data bases if charges were increased.

As of May 1982, the effects of the new pricing structure of on-line services for the *users* that access MEDLARS were unknown. There are a variety of user institutions with a large variety of charging procedures for on-line access. Most pass the costs through to the users in varying degrees, but others absorb the costs. It is reasonable to suppose that commercial firms would not be unduly affected by higher costs of on-line access even if creation costs were to be included in the full cost formula that includes creation costs. Even some of the major academic health center libraries might financially accommodate to the approximate additional cost of $1 or $2 per search which would result from the added costs.

But there is valid concern that students, nurses, and allied health personnel and researchers who are not working on grants or contracts might find it difficult to pay much more than NLM currently charges. The libraries in small hospitals might be adversely affected even by small increases in charges. Indeed, there is anecdotal evidence that there has been a 30-percent decrease in searches performed in a Northeast consortium of small-sized hospitals since the increase to $15 to $22 per connect hour by NLM (50). Smaller institutions are extremely limited in their ability to control costs, and the information center or the library is one of the few areas subject to cost control.

Differential Pricing

The price to access NLM data bases might be prohibitive to some users either if the full costs of accessing the system, or if the full costs of accessing the system and creating the data bases were recovered in on-line charges. The Government could decide to continue to provide access to all users by subsidizing the system. Other Government alternatives would be to set different charges for accessing MEDLARS data bases online for different types of users (i.e., differential pricing) or offer a subsidy (through grants for example) to particular institutions or individuals to allow them to select the service and data base they wish to use.

If the Government subsidized the system, theoretically, then as now, MEDLARS users would not have a choice of information services and would be dependent on NLM, which might or might not be responsive to their needs. This assumption about users' influence holds if NLM is providing on-line access to MEDLARS at a lower price than commercial information services and if price is the only, or even the main, factor in the choice of services. If the Government subsidizes the system another way—i.e., by paying private information services to provide the information at a low rate—users also would have little influence on the system. They would find it difficult to withhold payment or to switch services if they were dissatisfied.

Differential pricing is the alternative to governmental subsidization of MEDLARS that is currently receiving the greatest attention. Many proponents of differential pricing feel that the Government's subsidization of MEDLARS results in NLM's having an unfair competitive advantage over private information organizations. Some proponents also think it unfair that for-profit organizations pay the same on-line access charges as not-for-profit organizations to access the NLM data bases on MEDLARS. Differential pricing might eliminate or modify NLM's supposed unfair advantage and promote the information activities of the private information sector. The Government might also receive more revenues from increasing NLM's on-line access charges to commercial firms and other for-profit organizations.

But there are a number of arguments against differential pricing. NLM has been opposed to it in the belief that all users should have equal access to NLM services and all sectors of the user community should be charged the same amount for NLM's products and services. "Domestic commercial enterprises presumably pay their legislated share for supporting Federal Government activities and should enjoy the fruits of those activities on an equal footing with the non-profit sector" (110). In addition, establishing different charges for different institutions may require the imposition of a means test, which is not only demeaning to those who have to prove their inability to pay, but is usually cumbersome and costly to administer.

A problem in differential pricing is determining the category or categories of users eligible for lower prices. One could argue that not-for-profit organizations fall into this category. The Federal Government has historically recognized a distinction between not-for-profit and for-profit organizations. This distinction recognizes the important role played by nonprofit organizations in America. It recognizes the contribution which the private nonprofit sector has made towards achievement of social goals.

Precedent for the distinction between treatment by the Government of profit and nonprofit organizations can be found in many laws and policies. For example, the Internal Revenue Code recognizes, for tax purposes, a fundamental distinction between the private, nonprofit corporation organized under 501(C)(3) of the code, and the for-profit corporation. One pays taxes, and one does not. The code also recognizes that a deduction can be made for contributions and bequests by individuals to private, nonprofit organizations. Such contributions are not taxable. But one could also argue that not all not-for-profit organizations benefit the public and that some for-profit firms provide greater social benefits. In addition, the distinction between not-for-profit organizations is increasingly blurred. For example, the National Institutes of Health, which once limited its grants to not-for-profit organizations, now provides grants to for-profit firms as well.

The other alternative to governmental subsidization of MEDLARS is to subsidize MEDLARS users by grants. This may also require adherence to some type of qualifying criteria ("means test"). The key disadvantages of subsidizing users in this manner are the complexity and the costs both of making payments to hundreds of users and of enforcing the regulation. Furthermore, the potential for Government regulation of private information firms exists if the Government provides funds that could be used to purchase on-line access from private information services. Indeed, a Government subsidy using a mechanism similar to grants could be structured even if NLM were not providing on-line access to MEDLINE and its other data bases.

However, there are advantages in subsidizing some users of MEDLARS rather than subsidizing the system. Subsidizing users might encourage competition and stimulate private sector information activities, and might increase the efficiency of on-line information systems. Also, if the administrative costs do not exceed the charges recovered, the Government would recover more from user charges for on-line services if only some, as opposed to all, users did not pay the full costs for the services.

One observer has suggested that an office within the Department of Health and Human Services could offer credits for bibliographic search services of up to a given amount, such as $300 to applicants without other sources of funds (43). If the applicants stated that they were not working on sponsored research and that they had incomes and net worths below specified minimums, they would be granted the credits. A somewhat looser procedure could be followed by schools without asking for income and net worth, much as computing funds are now allocated in many schools, depending on the total number of students in the categories of interests. The above techniques should be capable of being administered at a cost of perhaps 10 to 20 percent of the subsidies granted. Naturally, more assurance of the legitimacy of the requests could be obtained, but only at higher cost (43).

Research and Development

NLM conducts, supports, and promotes basic and applied research in information science and its technologies. Research conducted at the Library's Lister Hill National Center for Biomedical Communications and under research grants and contracts, along with the efforts of many other governmental bodies, played a major role in the formation and development of the computerized data base and on-line information industry. Subsequently, private enterprise developed more advanced and innovative technologies (see chs. 2 and 4).

Research conducted and sponsored by NLM continues to benefit the private information sector. Among its other contributions, this research has laid the groundwork for the private production of master tapes for the subsequent production of video disks and the private development of video disk files of graphical data from patents, and has been responsible for the establishment of many private information firms (165).

NLM has also been among the first to recognize the need of practicing health professionals for more direct access to biomedical information than bibliographic sources afford. Thus, need is particularly acute in areas where primary information is limited or unavailable, as in developing countries. NLM developed the Hepatitis Knowledge Base as a "prototype information system" to enable users to quickly and efficiently find the proper information for their needs. Because of the intellectual input (the contents of the data base are evaluated by a consensus of experts), it is very expensive to produce. The Hepatitis Knowledge Base may serve only as a model of a refereed data base. Indeed, the American Medical Association and General Telephone & Electronics Corp. are launching a knowledge base containing drug-related information this year, but it will be of lesser magnitude and sophistication than the Hepatitis Knowledge Base and is expected to be used in concert with bibliographic data bases (128).

FINAL COMMENTS

This chapter has presented arguments concerning NLM's creation of computerized health-related data bases, its provision of on-line access to data bases, and different strategies for pricing the data base tapes and on-line access to the base. The discussion has focused on MEDLINE, the original and major data base in the Library's computerized retrieval and technical processing system, MEDLARS. The findings with respect to the issues follow.

OTA finds that over the years NLM has established strong and intimate ties with health and information communities worldwide who rely on the Library's information sources, and, in turn, contribute to the high quality and comprehensiveness of NLM's bibliographic sources.

A major finding is that the creation of MEDLINE by the Library seems to be warranted by NLM's extensive collection of biomedical materials, by its legislative mandate, and on economic grounds. NLM has the world's largest collection of biomedical literature and is mandated by law to organize and make its acquisitions available. For more than a century, *Index Medicus* has been the guide to the Library's collection of biomedical journals, and today the data base tape that is used in printing *Index Medicus* is also used in preparing MEDLINE. MEDLINE is also part of the Government's investment in biomedical research and assures ongoing access by researchers and practitioners to information needed to maintain and improve the public's health.

In addition, OTA finds that the possibility that the private sector would be inclined to create MEDLINE if NLM were to cease doing the activity is a matter of speculation. The information industry is young: it has been functioning for approximately 15 years. It is uncertain that if NLM were to cease creating MEDLINE a new or established private firm would have the desire to produce a similar product. As noted previously in this chapter, the other major health-related data bases complement rather than duplicate MEDLINE, and the availability of diverse bases is advantageous for the user.

Another finding is that there is no convincing argument that clearly supports any specific method of setting leasing fees for the MEDLINE data base tape. There is a wide range of interests between the public and private sector and within the private sector resulting in equally good reasons for leasing the data tapes to domestic firms at the cost of reproduction or at the costs of reproduction and creation. The economic arguments for instituting differential leasing fees are opposed on technical, scientific, and international grounds. On balance, arguments for changing the present policy are not convincing.

OTA finds no compelling reasons at present for NLM either to continue or to discontinue providing on-line access to MEDLINE. NLM has nearly achieved recovering the full costs of accessing the system, thus making its on-line charges more in line with, although still lower than, the charges of commercial information services. NLM's current on-line charges appear to be sufficiently low to adversely affect the ability of one of the two commercial vendors of MEDLINE to realize a profit in providing the same service. At the same time, preliminary evidence indicates that the charges are sufficiently high to prevent some small hospitals from continuing the same level of searching they previously performed on MEDLINE using NLM's system.

Another finding is that there is inconclusive evidence with which to weigh the advantages of charging the users of MEDLINE who cannot pay NLM's current on-line rates (or any future increase in costs) a lower rate than those who can pay such charges against the advantages of keeping the charges at a level all users can pay.

In summary, OTA finds that many of the arguments presented by proponents or opponents on the issues pertaining to the creation, provision, and pricing of NLM products and services seem more reflective of philosophical perspectives than objective analysis, and there are few empirical data to support them. Thus, any changes in the range of NLM's computerized products or services or in their pricing structure require caution.

OTA also finds that the rapidly changing nature of the computer and communications fields gives additional credence to the need for care in current information policies and practices. The information field is in a period of flux: the shape of current on-line information systems is expected to change within a few years, particularly in the areas of remote processing and software. Along with technological changes, the economic issues pertaining to information systems, including MEDLARS, may change. *For this reason, OTA concludes that decisions made today in reaction to current problems should be, to the maximum extent feasible, informed by the ways that technological advances might change those very issues. OTA urges that this report be considered in the context of the material presented in appendix H on future information technologies.*

Appendixes

Appendix A.—The National Library of Medicine: Organization and Activities*

Introduction

The National Library of Medicine's (NLM or the Library) organizational structure and intramural activities primarily reflect the legislative intentions of the National Library of Medicine Act of 1956 (Public Law 84-941). Similarly, its extramural programs are grounded in the Medical Library Assistance Act of 1965 (Public Law 89-241). These congressional actions serve to organize the discussion that follows. Three other legislative and executive mandates have also shaped the Library and, while not discussed in detail, warrant mention: the transfer of the Public Health Audiovisual Center (now the National Medical Audiovisual Center) to NLM in 1967; the development of the Toxicology Information Program, also in 1967; and the founding of the Lister Hill National Center for Biomedical Communications in 1968.

National Library of Medicine Act of 1956 (Public Law 84-941)

NLM began as a small collection of medical books and journals in the Office of the Army Surgeon General in 1836. By 1895, it was international in scope, and had grown from some 1,800 volumes to over 117,000 books and 192,000 pamphlets covering almost every medical topic. In 1922, it was named the Army Medical Library. In 1952, it was renamed the Armed Forces Medical Library to reflect its broadening user community. In 1956, with passage of the National Library of Medicine Act (Public Law 84-941), the collection was recognized as a "great National medical resource" serving the Nation's entire medical community.

Congressional Intent

Despite its lofty position as the "largest and most important medical library in the world," by 1956 the Armed Forces Medical Library was beginning to suffer under an awkward administrative arrangement, inadequate for the increasingly diversified demands levied against its resources. In hearings before the House Committee on Interstate and Foreign Commerce, the Library was reported to be "inadequately housed in a building where its collections are threatened by loss from fire and by damage through exposure to the weather." Administered by the Department of Defense, its funding had been subject to wide fluctuations. The committee concluded that "difficulties of operation have arisen because no clear authority exists for many of the functions which the Library now serves."[1]

Through the National Library of Medicine Act, Congress sought to assist "the advancement of medicine and the related sciences, and to aid the dissemination and exchange of scientific and other information important to the progress of medicine and to the public health" (Public Law 84-941). It wished to improve the health of people in the United States by providing access to information for health professionals and policymakers. Congress did not assign the provision of health information to the public to NLM, but turned to other branches of the Public Health Service to carry out this responsibility.

Authorizing Legislation

An amendment to the Public Health Service Act, the National Library of Medicine Act established NLM in 1956, and authorized it to acquire, preserve, and make available materials pertinent to medicine; to prepare and make available indexes, catalogs, and bibliographies of the materials; and to provide reference and research assistance. The act established a Board of Regents whose members, appointed by the President, are to advise the Surgeon General on "important matters of policy in regard to the Library." Appropriations for the construction of facilities adequate for the Library's use were also authorized. Finally, the act transferred the Armed Forces Medical Library from the Department of Defense to the Public Health Service.

Congress recognized the Library's importance to the advancement of medicine in both the United States and throughout the world. Placed under the auspices of the Public Health Service, the Library is in a position "where contact with and participation in programs of medical research will provide the best environment for the Library's proper functioning and continued growth."[2]

The National Library of Medicine Act does not require renewal, and since its enactment has had few changes. One important change was made in 1978, when the power to appoint members to the NLM Board of Regents was transferred from the President to the Secretary of Health, Education, and Welfare

---

*Information in this appendix was primarily obtained from the staff and publications of the National Library of Medicine in 1981.

[1] U.S. Congress, House of Representatives, House Report 84-941, 84th Cong., 1956.

[2] Ibid.

(Public Law 95-622). The amendment was prompted by the need to quickly fill vacant seats on the Board.

Organization of NLM

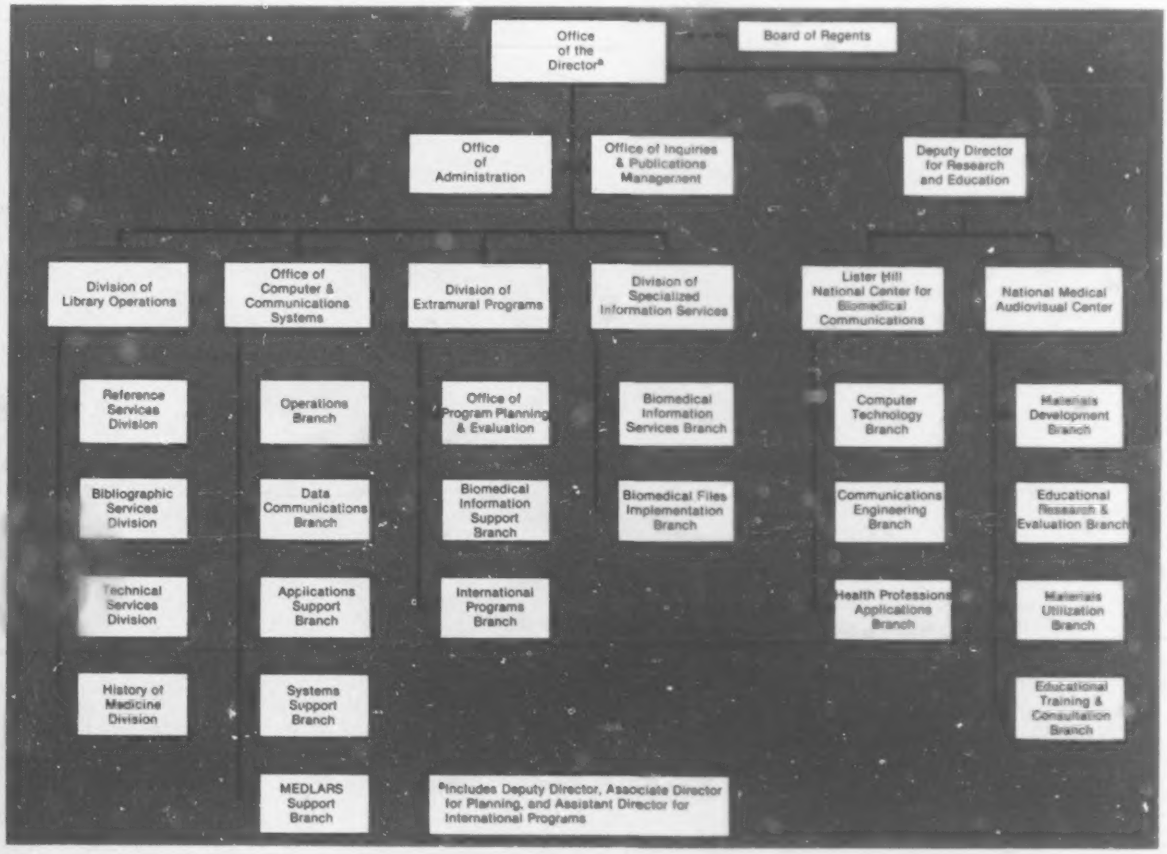
Since its designation as a national library in 1956, NLM has continually expanded its scope of operations and responsibilities. As evident from its current organizational structure (see fig. A-1), the Library has had many functions added to those authorized in the original act. The organizational structure reflects both legislative and executive actions from 1956 to 1968. The last major addition to NLM was made in 1968, when the Lister Hill National Center for Biomedical Communications was established as the Library's internal research arm. That same year, NLM was transferred from the Surgeon General's Office to the National Institutes of Health (NIH).

BOARD OF REGENTS

In accordance with legislative requirements, the 10 members of the NLM Board of Regents, appointed by the Secretary of Health and Human Services, are leaders in the following disciplines: the fundamental sciences, medicine, dentistry, public health, hospital administration, pharmacology, scientific or medical library work, public affairs, and representatives of the general public. The Regents serve overlapping 4-year terms. There are also seven ex officio members: the Surgeons General of the Public Health Service and the three Armed Services; the Chief Medical Director of the Veterans Administration; the Assistant Director for Biological, Behavioral, and Social Sciences of the National Science Foundation; and the Librarian of Congress.

Historically, the disciplines Board members represent have generally adhered to legislative requirements.

Figure A-1.—National Library of Medicine


Office of the Director[a]
Board of Regents
Office of Administration
Office of Inquiries & Publications Management
Deputy Director for Research and Education
Division of Library Operations
Office of Computer & Communications Systems
Division of Extramural Programs
Division of Specialized Information Services
Lister Hill National Center for Biomedical Communications
National Medical Audiovisual Center
Reference Services Division
Bibliographic Services Division
Technical Services Division
History of Medicine Division
Operations Branch
Data Communications Branch
Applications Support Branch
Systems Support Branch
MEDLARS Support Branch
Office of Program Planning & Evaluation
Biomedical Information Support Branch
International Programs Branch
Biomedical Information Services Branch
Biomedical Files Implementation Branch
Computer Technology Branch
Communications Engineering Branch
Health Professions Applications Branch
Materials Development Branch
Educational Research & Evaluation Branch
Materials Utilization Branch
Educational Training & Consultation Branch
[a]Includes Deputy Director, Associate Director for Planning, and Assistant Director for International Programs

SOURCE: National Library of Medicine, January 1981.

BEST AVAILABLE COPY

However, the Secretary has some latitude in his appointments, so they tend to reflect the more immediate interests of the Library. For example, in fiscal year 1981, there were no obvious appointments from the public health field or the health services research community, although there have been such members in the past. Rather, two Board members had expertise in computer systems or the computerization of biomedical information, as NLM is currently concentrating many of its resources on the development of MEDLARS III, the next generation of its computerized on-line bibliographic retrieval system.

The Board meets three times a year and provides advice to the Secretary of Health and Human Services, the Assistant Secretary for Health, the Director of NIH, and the Director of NLM on Library policy. The Board makes recommendations on "the acquisition of materials for the Library, the scope, content and organization of the Library's services, and the rules under which its materials, publications, facilities, and services shall be made available to various kinds of users" (Public Law 84-941). The Director of NLM has the responsibility for operating the Library and so, in essence, the Board is advisory to him. The Board promulgates policy, and is the final review body for extramural grant applications, which are evaluated for program and policy relevance.

ORGANIZATIONAL DIVISIONS

Division of Library Operations.—The Division of Library Operations is the traditional core of the Library. It performs NLM's basic activities—collecting, organizing, indexing, cataloging, and making available much of the world's biomedical literature—and has been instrumental in creating and adopting new techniques to improve library services. Its four subdivisions and their functions are listed below:

- The Reference Services Division processes interlibrary loans, provides reference and bibliographic services, and maintains and preserves the Library's general collection.
- The Bibliographic Services Division indexes serial literature for *Index Medicus* and other indexes, enters references into the data bases, and coordinates the on-line network that makes references available via NLM's on-line data bases such as MEDLINE.
- The Technical Services Division selects and acquires printed material; catalogs and classifies books, monographs, Government documents, and other materials; and makes the information available through publications and on-line for other libraries to use. It is the national authority for bibliographic control of biomedical publications.
- The History of Medicine Division acquires, processes, and makes publicly available the Library's collection of historical biomedical books and journals. It also maintains active public relations and research programs on the history of medicine.

Office of Computer and Communications Systems.—This office provides data processing and communications support for all NLM operations. In addition to its routine activities, the office works toward improving the performance of NLM's computer equipment, adding new technological features and enhancing the capabilities of MEDLARS, and is developing a data communications system to manage the internal operations of the Library. As part of NLM's efforts to reach out to the library community, the office developed an innovative information management system that supported the 1980 White House Conference on Library and Information Services.

Division of Extramural Programs.—The Division of Extramural Programs administers five of the six grant programs that are authorized by the Medical Library Assistance Act of 1965 (Public Law 89-241). The sixth is the Regional Medical Library Program, which is administered from the Office of the Associate Director for Library Operations. All six programs are discussed in this appendix.

Division of Specialized Information Services.—The diversity of NLM's responsibilities is particularly visible in the Division of Specialized Information Services. The division operates the Toxicology Information Program, established at NLM in 1968 to centralize access to information on toxicology. The program draws toxicology information from several Federal and private, nonprofit agencies. The division has established computer-based toxicology data banks from information in the scientific literature and from the files of collaborating industrial, academic, and governmental organizations. It also established and administers toxicology information services for the scientific community.

The division's other activities include the Toxicology Information Response Center in Oak Ridge, Tenn., which performs literature searches in toxicology and environmental health; produces a number of publications, including one on toxicology testing and one on toxicology research; and conducts a series of collaborative projects with other Government agencies.

Lister Hill National Center for Biomedical Communications.—The center is responsible for developing methods to improve information transmission so that health professionals will have easier access to the ever-growing volume of information. It is the research and development branch of the Library and investigates the use of computers and communications technology in advancing health education, biomedical research, and health care delivery. Since the dedication of its new facility, the Lister Hill Center Building, in May 1980, the center has emphasized intramural research.

One of the center's first accomplishments was assisting in the development of NLM's on-line retrieval system. It has also conducted research programs on the effectiveness of orbiting satellites for communicating medical information, the use of two-way television using microwave links for continuing medical education, and the use of computer-assisted instruction. Current and future projects include the development of knowledge-based programs in specialized areas of biomedicine that will make new medical findings and research information rapidly available to health professionals, particularly practitioners; and the design, development, and evaluation of an experimental storage and retrieval system to electronically scan, store, retrieve, and display documents acquired by NLM.

National Medical Audiovisual Center (NMAC).— The aim of NMAC is to improve the quality and use of biomedical audiovisuals in health professional schools and the biomedical community. Before it was transferred to NLM in 1967, NMAC was a component of the Center for Communicable Diseases (now Centers for Disease Control, CDC), and, in fact, remained in its Atlanta facilities until March 1980. As part of CDC, the center produced films of award-winning quality for the health education of high school students and the public. When it became part of NLM, the Board of Regents reoriented it to conform to the Library's legislative purpose, i.e., providing health professionals access to information.

NMAC's current activities include research and evaluation in audiovisual design and medical photography, training health educators in the use of audiovisual technology, and the management and distribution of a large collection of medical motion pictures and videotapes. NMAC was the original developer of AVLINE, an audiovisual data base, which was transferred to the Division of Library Operations in 1977. NLM plans to emphasize NMAC's research function in the future, and merge NMAC with the Lister Hill Center (102).

INTERNATIONAL ACTIVITIES

NLM has been active in international programs since John Shaw Billings became the Librarian of the Library of the Army's Surgeon General in 1865. Wanting to develop a collection that was international in character, Billings collected library materials from throughout the world and began an exchange program with foreign libraries, medical schools, and other scientific institutions. Today, NLM has formal exchange agreements with 382 institutions in 72 countries.

NLM's Special Foreign Currency Program, authorized under the Agricultural and Trade Assistance Act of 1954 (Public Law 83-480), supports the preparation of secondary literature, including critical reviews by outstanding scientists in particular fields, and translations of foreign monographs in the health sciences. The program is currently active in Poland, Egypt, Tunisia, India, Yugoslavia, and Pakistan, and in Israel under awards from the U.S.-Israel Binational Science Foundation. Other international programs include the exchange of biomedical literature, the provision of library services such as interlibrary loans to foreign institutions, the specialized training of qualified individuals from abroad who have national or international sponsorship, technical consultation and collaboration with governmental and nongovernmental international organizations, and participation, as appropriate, in formal U.S. bilateral health agreements.

There are, as well, bilateral MEDLARS agreements with 13 foreign countries—Australia, Canada, Colombia, France, Italy, Japan, Kuwait, Mexico, South Africa, Sweden, Switzerland, United Kingdom, and West Germany—and with the Pan American Health Organization. These agreements do not require the direct expenditure of U.S. moneys, nor the expenditure of foreign funds to the United States, but provide centers in these countries with access to MEDLARS in exchange for indexing services that foreign centers either perform on a quid pro quo basis or pay U.S. commercial contractors to perform. Operational decisions, such as determining who accesses the data bases, are the responsibility of the foreign center.

Appropriations and Staffing

NLM's appropriations in 1970 to 1982 are displayed in table A-1. A continuing resolution appropriated $44.4 million for 1982. Funds for the extramural grant program declined by more than $2 million from the 1981 level, while funds for library operations increased by more than that amount. NLM will use most of the operations increase for development of MEDLARS III. The decrease in the grant program was required in the reauthorization of the Medical Library Assistance Act for 1982.

The full-time permanent staffing level remained fairly stable from 1975 to 1980, reaching a high of 495 in 1978. However, in recent years part-time temporary staff have swelled the ranks (see table A-2). During these same years, the services rendered by NLM have increased substantially (see fig. A-2).

Medical Library Assistance Act of 1965 (Public Law 89-241)

After World War II, the Federal Government assumed an increasing responsibility for funding scientif-

Table A-1.—NLM Appropriations, 1970-82 (dollars in thousands)

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982
Medical Library Assistance Act													
Grants	$3,950	4,080	4,790	4,547	4,455	4,330	3,712	5,375	5,187	5,987	6,725	6,831	5,000
Contracts	1,842	1,912	2,102	2,075	2,574	2,352	2,721	2,625	2,800	3,000	3,200	3,000	2,500
	$5,792	5,992	6,892	6,632	7,029	6,682	6,433	8,000	7,987	8,987	9,928	9,831	7,500
Intramural													
Lister Hill Center	$945	1,456	1,960	2,055	2,103	2,863	2,922	4,952	5,031	6,255	5,554	5,105	5,045
National Medical A/V Center	2,224	2,196	2,556	2,795	2,888	3,263	3,303	3,846	4,074	4,343	4,350	4,198	3,125
Library Operations	7,431	8,348	9,030	9,018	9,502	10,860	10,821	12,255	13,147	15,258	16,014	17,527	20,082
Toxicology Information	1,552	1,280	1,370	1,590	1,627	1,915	1,947	2,277	2,334	2,401	3,368	3,105	3,299
Direct Operations	594	810	985	1,160	1,236	1,389	1,582	1,563	1,661	1,728	1,808	1,788	1,821
Program Management	1,144	1,354	1,332	1,910	1,944	1,874	2,057	2,341	2,512	2,459	2,981	3,092	3,330
Total	$19,682	21,436	24,127	25,150	26,329	28,850	29,065	35,234	36,746	41,431	44,000	44,666	44,402
Total (in constant dollars)	$19,682	20,280	21,815	21,812	21,688	21,840	20,850	23,823	23,024	24,102	23,822	22,673	21,660

SOURCE: National Library of Medicine.

Table A-2.—NLM Staff, Fiscal Years 1975-81
(actual on-board count)

Activity		1975	1976	1977	1978	1979	1980	1981
Lister Hill Center	P	22	24	24	35	30	34	40
	O					3	5	9
National Medical A/V Center	P	101	101	88	88	76	37	56
	O					7	1	4
Library Operations	P	196	201	212	221	212	202	217
	O					38	36	24
Office of Computer Services	P	52	54	52	51	51	51	54
	O					4	4	3
Specialized Information Services	P	17	18	17	18	19	19	27
	O					3	4	4
Extramural Programs	P	22	24	27	25	25	23	22
	O					6	5	8
Program Direction	P	48	50	52	57	55	62	58
	O					11	9	14
Total	P	458	472	472	495	468	428	474
	O					72	64	66
		458	472	472	495	540	492	540

P - Permanent full-time.
O - Other.

SOURCE: National Library of Medicine.

ic research, particularly in the health sciences. The result was an explosion of new scientific information in the biomedical and health sciences that overwhelmed the ability and facilities of health science libraries. Libraries had received little Federal support comparable to that given research, and lacked the trained personnel, resources, and techniques to ensure that the full value of this new knowledge could be realized. By 1965, the growth of information had also outstripped health practitioners' ability to apply it, prompting the need for continuing education programs to update professionals' skills. The possibility of using medical libraries as educational tools for health professionals was an important factor in enacting the Medical Library Assistance Act of 1965 (3).

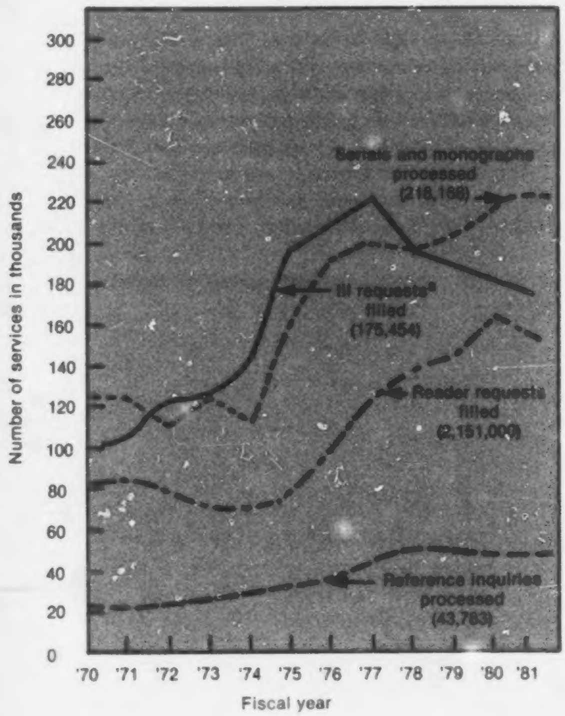
Congressional Intent

In hearings before the House Committee on Interstate and Foreign Commerce in 1965,[3] studies were cited indicating a need for additional library space, more trained biomedical librarians, additional volumes and periodicals, support for research and development projects in biomedical communications and medical

---

[3] U.S. Congress, House Report 89-1026 accompanying H.R. 3142, Interstate and Foreign Commerce Committee, Sept. 20, 1965.

Figure A-2.—Representative NLM Services, 1970-81



[a]Interlibrary loan.

SOURCE: National Library of Medicine.

library science, and a system of regional medical resource libraries to ensure access to medical documents and avoid the costly duplication of extensive collections.

Congress thus faced a series of interrelated problems. Considerable Federal assistance had been directed toward "the intensive development of health research institutions, medical schools, and other medical facilities," including hospitals, and to increasing the supply of physicians, nurses, and other health professionals.[4] Concurrently, and again largely due to Federal funding, the knowledge bases of medical research and practice were broadening, drawing on many new fields and disciplines, and requiring, as well as adding to, a widening, complex body of literature.

During this period, there was little support for medical libraries. The condition of many deteriorated from the pressures of rapidly developing programs in medical institutions, and they were unable to meet the demands of a dynamic medical community (37). In 1964, a Presidential commission noted "that unless major attention is directed to improvement of our national medical library base, the continued and accelerated generation of scientific knowledge will become increasingly an exercise in futility" (121).

At the urging of numerous professional and academic associations, Congress enacted the Medical Library Assistance Act of 1965 (Public Law 89-241) to address medical libraries' needs for additional resources and personnel to meet the demands of a growing user population and an expanding body of biomedical and health science information. Through the act, Congress hoped to strengthen local and regional health science libraries so that researchers and practitioners could keep more fully informed of research findings and new medical developments, and ultimately provide better health care for the American people.

Authorizing Legislation

The Medical Library Assistance Act reflected a fundamental change in Government policy regarding support for libraries allied with the health sciences. An amendment to the Public Health Service Act, it provided financial assistance for the development of facilities and techniques necessary to "collect, preserve, store, process, retrieve, and facilitate the dissemination and utilization of biomedical and health science . . . knowledge and information" (Public Law 89-241).

The act outlined a 7-point approach to strengthen the Nation's medical libraries. The Surgeon General, through NLM, was authorized to:

- assist in construction of new and renovation of old medical library facilities;
- assist in the training of medical library personnel and personnel in fields related to health;
- financially assist physicians and other scientists in the compilation of existing and new scientific knowledge;
- assist in the development of innovative technological advancements in medical library techniques;
- assist in the expansion of the resources and services of medical libraries;
- assist in the establishment of a system of regional medical libraries to coordinate the geographic sharing of resources; and
- assist financially in the publication of biomedical science works.

The Medical Library Assistance Act of 1965 also established the authority to assist in the establishment of regional branches of NLM in the National Library of Medicine Act of 1956.

NLM had previously submitted almost identical legislative specifications to Congress and the administration. One of its concerns was that, with the continued dependency of local libraries, NLM would evolve into a "monolithic medical resource in this nation" (37). NLM believed that the country required "the development of a complex of regional medical libraries . . . with adequate facilities, resources, and personnel to serve those sections of the Nation with underdeveloped library facilities" (37).

---

[4] Ibid.

The act authorized NLM to provide financial assistance, through a system of competitively awarded grants and contracts, to "all appropriate public and private institutions and individuals active in the provision of health services or in health-related teaching and research" (Public Law 89-241).

Congress has reauthorized the Medical Library Assistance Act six times since 1965. Although the original intent remains, the legislators have modified some portions of the act over the years. For instance, in the original legislation, Congress intended "medical libraries" to be defined in the broadest sense, to include all libraries affiliated with health and biomedical sciences, though the need for this legislation was largely defined in terms of the needs of medical school libraries. In 1970, the title of the act was changed to include "health communications," adding emphasis to the breadth of intent for participation in the programs. Eligibility extended to all "clincial fields including medicine, dentistry, optometry, pharmacy, osteopathy, veterinary medicine where relevant to human health, nursing, public health, other health-related fields, and fundamental and applied sciences when related thereto" (Public Law 91-212).

In 1973, Congress removed the authority to assist in the construction of library facilities. The next year, it authorized a single appropriation for the assistance programs, leaving the allocation of funds for individual programs to the discretion of the Library. In the 1978 reauthorization, NLM was encouraged to "play a more aggressive role in the collection and dissemination of research findings directly relevant to clinical practice" and was reauthorized through September 1981. The most recent reauthorization extends the Medical Library Assistance Act through September 1985.

Appropriations

The 1965 Medical Library Assistance Act authorized $105 million for 1965 to 1970 for NLM to initiate programs assisting the Nation's medical libraries and the health science community. However, only $40.8 million was appropriated for this period. In assessing the achievements of programs established under the act, the Director of the Library concluded that although NLM had significantly improved medical libraries and information resources and services, the objective of the act had not been fully realized due to insufficient funding (37). This sentiment was echoed in the Senate Report accompanying the 1970 reauthorization.[5]

As can be seen in table A-3, a large discrepancy between the funds authorized and those appropriated persisted until the most recent reauthorization. For fiscal year 1981, $9.8 million was appropriated, and $7.5 million is scheduled for fiscal year 1982. Although there is an apparent growth of appropriated funds from 1970 to 1981, figure A-3 indicates an actual decrease in constant dollars.

Extramural Grant Programs

NLM's Division of Extramural Programs originally administered seven authorities, but, as noted earlier, Congress removed the authority to assist in construct-

Table A-3.—Medical Library Assistance Act Authorizations and Appropriations, Fiscal Years 1975-82 (dollars in thousands)

Fiscal year	Funds authorized	Funds appropriated	Percentage of authorized funds appropriated	Difference between authorized and appropriated funds
1975	$17,500	$6,682	38	$10,818
1976	20,000	7,658[a]	38	12,697
1977	20,000	8,983[b]	62	5,613
1979	15,000	8,986	60	6,014
1980	16,500	9,924	60	6,576
1981	18,500	9,831	53	8,669
1982	7,500[c]	7,500	100	0

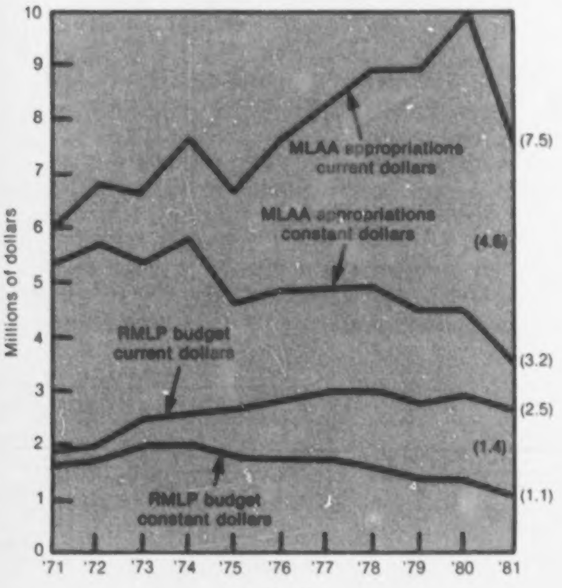
[a] Includes fiscal year transition quarter, July 1, 1976-Sept. 30, 1976.
[b] Includes original appropriation of $7,987 thousands and an additional $1 million of reprogramed funds approved by Congress.
[c] Authorization extends through September 1982.

SOURCE: National Library of Medicine.

---

[5] U.S. Congress, Senate Report 91-480 accompanying H.R. 11702, Labor and Public Welfare Committee, Oct. 16, 1969.

Figure A-3.—Medical Library Assistance Act (MLAA) Appropriations and Regional Medical Library Program (RMLP) Budget Current and Constant Dollars, Fiscal Years 1971-81
(based on 1969 constant dollars)



SOURCE: National Library of Medicine.

ing library facilities from the Medical Library Assistance Act in 1973. The authority had been exercised only between 1968 and 1969, when NLM awarded grants to nine medical schools, one school of optometry, and one school of veterinary medicine with the $11.5 million appropriation.

In 1981, the Associate Director for Library Operations was assigned responsibility for the Regional Medical Library Program (RMLP). Nevertheless, all the extramural programs are interrelated, and many of the grant programs administered in the Division of Extramural Programs, particularly the resource grants, promote the aims of RMLP. (RMLP is discussed below.)

The division sees its role as interpreting and advancing the intent of Congress by means of various grant mechanisms. The major areas of emphasis have altered over the past 15 years in response to the health professional community's perceived needs. Currently, the grant programs emphasize biomedical communication, e.g., the storage and retrieval of biomedical information, the role of computers in medicine, and librarianship. The grant mechanisms the division uses are:

1. resource grants;
2. training grants in health sciences and computer technology;
3. research, development and demonstration grants;
4. special scientific projects; and
5. publication grants.

Each of the grant programs is described below, including its evolution from 1965 to the present, its current status, and OTA's assessment with respect to congressional objectives. Table A-4 displays the distribution of funds among the grant programs for fiscal years 1980, 1981, and 1982.

RESOURCE GRANT PROGRAM

At its inception, the resource grant program was primarily intended to correct deficiencies in collections, equipment, and organization of collections of established libraries. Health science libraries' resource needs appeared to be endless: in 1965, Congress observed that the libraries needed more than $85 million, and authorized $5 million per year for a 5-year period. Only $11.8 million were appropriated for those 5-years however; NLM supported 401 libraries rather than its goal of 600 to 700.

Distribution of the funds from 1965 to 1970 reflected the formula grant then in use. The grant was based on a library's budget, and as a result, larger libraries received a larger proportion of the funds. Medical libraries in academic settings received 23 percent of the grants awarded, but more than 62 percent of the funds. The smaller hospital libraries received more than 50 percent of the grants awarded, but only 26 percent of the distributed funds. Other awards went to schools of dentistry, pharmacy, and veterinary medicine, as well as other academic institutions, State institutions, and professional societies. Grants could be used to acquire library materials, increase staff, or purchase materials.

Because more than 50 percent of the funds was used to increase libraries' collections, the Director of NLM concluded that the most immediate need had been the acquisition of books, journals, and other publications,

Table A-4.—Medical Library Assistance Act: Distribution of Funds Among Grant Programs, Fiscal Years 1980-82
(dollars in thousands)

Program	1980	1981	1982
Resource	$1,596	$1,641	$1,206
Training	1,638	1,308	1,000
Research	2,724	2,774	2,257
Special scientific projects	142	290	23
Publications	787	818	514
Total	6,887	6,831	5,000

SOURCE: National Library of Medicine.

but that "the time had come for putting emphasis on improved service" (37).

In the 1970 reauthorization, Congress acknowledged that the grant program had been effective but that additional funding was required to bring health science libraries up to desired standards. In authorizing increased funding, Congress dropped the requirement for formula grants, but retained the provision of the original legislation that no institution receive more than $200,000 in grants in any fiscal year.

After assessing the first 5 years of the program, NLM decided that two types of grants were required. Most of the larger libraries had rebuilt their collections, due, in part, to the advantages offered by the formula grant. A Resource Project Grant Program was established so that existing *services* could be expanded or new ones created. The emphasis was on sharing resources and improving services (but not collections) by funding projects to streamline operations, utilize new technologies, and assist libraries in introducing and improving the use of new media, such as microfilm, audiovisuals, and computer-assisted instruction.

But it was also clear to NLM that many smaller (mainly hospital) libraries still needed assistance in developing their collections. In addition, the 1970 reauthorization had permitted grants to establish new libraries. Thus, in 1971 a Resource Improvement Grant Program was started that provided a 1-year, one-time grant award of up to $3,000 to assist in establishing and/or developing a basic information collection in smaller community hospital libraries.

As a result of an evaluation conducted in 1974 by the NLM Office of Program Planning and Evaluation, the Resource Improvement Grant Program was modified to fund consortium arrangements as well as individual institutions (84). Single institutions can be funded for 1 year to a maximum of $4,000, with a matching requirement of $1,000 from the institution to develop a collection. Grants are also available for up to 2 years to support activities necessary for the planning, organization, and development of a health science library consortium, which is composed of a number of libraries, usually within a defined geographic area, that agree to share resources.

Up to $4,000 can be provided per eligible consortium member, with a matching requirement of $1,000 to support the purchase of basic information collections. The resource improvement grant program objectives are to strengthen the Regional Medical Library Network by developing adequate health science library collections at the local level, and to encourage resource sharing among local health-related institutions. Resource improvement awards for individual institutions and consortia are considered "seed money" to further the program's purposes and are not to be used for operating expenses.

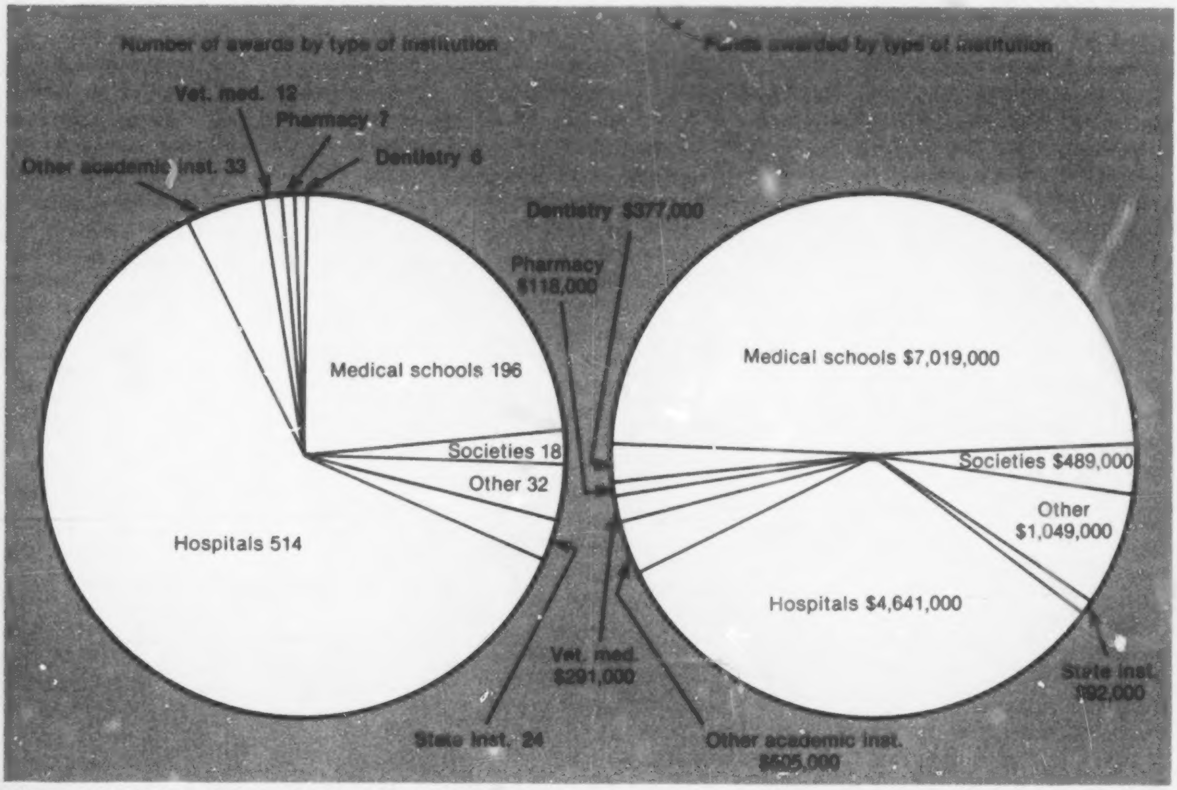
The consortium program has been well received by the health science library community. The program's goal was to organize 250 institutions in consortia within 5 years; it was realized in 2 years. At the outset of the program, an average of five institutions participated in each consortium. This average now stands at 10. Further, though Federal seed money is provided to each consortium for only 2 years, the size and number of consortia continue to grow. In part, this may reflect the increasing costs of maintaining collections and financing interlibrary loans, factors that make sharing resources more attractive. Currently, about half the Nation's 7,000 hospitals have access to an information facility, many through consortia arrangements.

As noted earlier, the formula mechanism used to determine the size of awards was an advantage for medical school libraries and a disadvantage for the smaller hospital libraries. The abolishment of the formula mechanism in 1970 partially corrected this discrepancy. Based on figures from NLM, libraries in medical schools received 48.5 percent of the funds awarded from 1971 to 1978, in contrast to 62 percent received from 1965 to 1970, and hospital libraries received 32 percent from 1971 to 1978, up from the 26 percent they had received from 1965 to 1970. Libraries in hospitals still receive the largest proportion of the awards (61 percent), while those in medical schools receive 23 percent (see fig. A-4).

In 1981, there were 74 active grants: 32 resource project grants and 42 resource improvement grants. Examples of the former are: 1) a Veterinary Medical Information Center Project; 2) an effort to organize the Adolf Meyer Papers; 3) a Computer Assisted Dental Simulation Project; and 4) a Community Information Network for Health Education Project. Funding for resource project grants varies widely, ranging from $8,000 to up to $452,000 for multiyear awards. Funds for the resource improvement grants develop consortia for better resource sharing or develop library collections; thus all the projects are similar in scope and title, as well as in funding: more than half receive from $3,000 to $4,000.

Table A-5 shows the number of awards and funding levels from 1971 to 1982. More than $1.5 million was allocated for fiscal year 1980 for 73 awards. The number of awards is considerably lower for fiscal year 1981, and there is a decrease in both funds and awards for fiscal year 1982: 42 awards totaling $1,206,000 are scheduled.

Figure A-4.—Distribution of Resource Grants, Fiscal Years 1971-78



SOURCE: National Library of Medicine, Extramural Programs.

Table A-5.—Medical Library Assistance Act: Resource Grants, Fiscal Years 1971-82 (dollars in thousands)

Year	Amount	Number of awards
1971	$2,231	469
1972	2,505	372
1973	2,298	153
1974	2,632[a,b]	127
1975	1,469	75
1976	726	43
1977	1,773	66
1978	2,013	52
1979	2,008	78
1980	1,596	73
Subtotal	19,251	1,509
1981	1,641	56
1982	1,206	42
Total	22,098	1,607

[a] Includes Regional Medical Library Grants.

[b] Excludes interagency agreement of $27,000 paid with carry-over funds from fiscal year 1973.

SOURCE: National Library of Medicine.

TRAINING PROGRAMS

Currently, Medical Library Assistance Act training grants support graduate level programs in computer technology within the health sciences. The objective is to promote the integration of computer technology with all phases of clinical medicine: teaching, practice, and research.

The focus of the training program has shifted considerably since it was initiated in 1965. Originally, it was to increase the number and quality of medical librarians. In the early 1960's, four distinguished advisory bodies concluded that there was a critical shortage of trained professionals to staff health science libraries and to meet the information needs of the health science community (78). Available resources were clearly not capable of alleviating the shortage; only 10 schools in the country offered even one course in medical librarianship, and just three medical libraries offered post master training programs, with a combined capacity of only eight places each year.

In 1965, Congress authorized $5 million to train 750 information personnel over a 5-year period. The appropriations were almost at the level of authorization: $4.5 million were used to establish 20 programs that trained 350 people in medical librarianship. In addition, eight fellowships were awarded to study the history of medicine and biomedical communication. At the end of 5 years, the Director of NLM reported that the program had met many of its goals, but that insufficient attention had been placed on the retraining of employed librarians in modern information handling techniques (37). The Extramural Program staff still considers this a problem, in that an expeditious way of retraining experienced professionals has not been found.

In 1970, Congress concluded that the program had had an encouraging beginning, but had not satisfied the identified need for medical librarians. Thus, it increased training grant funding at a higher level for the next 3 years. However, a 1973 NLM-funded study reported that the training grant program had perhaps produced too many medical librarians relative to the current job market. The study also noted that, at least in some sections of the country, the apparent shortage of the mid-1960's had been largely eliminated (51). Similarly, other studies released at about the same time corroborated the finding that by 1973 sufficient medical librarians were available to meet the needs of health science libraries. There remains, however, disagreement in the library community, especially in graduate programs, over the accuracy of these findings and conclusions.

NLM received these evaluations as indication of the training program's success. The staff of the Division of Extramural Programs note that the Library had not intended to fund long-term training programs, that the Library's justified expectation was that NLM funds would provide only the nucleus for growth, and that the programs would continue with support from academia. In fact, during the 1974 reauthorization hearings, Congress complimented the Library for accomplishing this objective.[6]

Aside from the supply of medical librarians, NLM's 1972 decision redirecting its program to train health professionals in the application of computer technology to medicine was prompted by a comprehensive report by the Association of American Medical Colleges (AAMC) on medical education technology.[7] AAMC found that major changes in the current system of libraries, publishing, and medical school curricula required personnel to be familiar with computer technology.

The Library initiated its training grant program in health sciences and computer technology after obtaining the approval of the appropriations committees of Congress. Currently, training grants are designed for health science faculty and potential faculty in the anticipation that their knowledge of computer techniques will be transmitted to the next generation of practicing physicians, researchers, educators, and other health professionals, and that computers will be utilized in solving medical problems. An evaluation of the program is now being designed at NLM.

Appropriations for the training programs are displayed in table A-6. Except for a dip in expenditures from 1973 to 1975, the table shows a gradual but steady increase in expenditures, uncorrected for inflation, from 1972 to 1980. In fiscal year 1980, of the $1.6 million available, more than half was to cover direct trainee expenses and the rest partially reimbursed the grantee institutions for additional expenses resulting from the training grants. The 1981 budget appropriates $1.3 million for the support of 10 training programs, and the 1982 budget has $1 million allocated for continuing nine of these programs.

For the past 3 years, the Library has funded a $343,000 experimental internship program for library administrators (82). This program was initiated in response to the frustrations of search committees unable to find individuals qualified to be directors of many large health science libraries. The librarians trained in NLM-sponsored and other programs had not yet attained the experience required to direct a large

Table A-6.—Medical Library Assistance Act: Training Grants, Fiscal Years 1971-82
(dollars in thousands)

Year	Amount	Number of awards
1971	$1,000	13
1972	1,234	15
1973	720	13
1974	901	11
1975	891	9
1976	1,389	13
1977	1,331[a]	11
1978	1,459[a]	11
1979	1,472[a]	10
1980	1,638	10
Subtotal	12,035	116
1981	1,308	10
1982	1,000	9
Total	14,343	135

[a] Includes Council of Library Resources Training Contract.

SOURCE: National Library of Medicine.

---

[6] U.S. Congress, Senate Report 93-764 accompanying H.R. 11385, Labor and Public Welfare Committee, Apr. 1, 1974.

[7] E. A. Stead, C. M. Smythe, C. G. Gunn, et. al. (eds.), "Educational Technology for Medicine: Role for the Lister Hill Center," *J. Med. Educ.* 46:1, 1971.

library. To date the NLM training program in library administration has produced nine graduates. When measured by the pragmatic criteria of employment, the program has had partial success, as most graduates have been hired as assistant directors or are under consideration for a directorship. NLM is now reassessing the program.

RESEARCH GRANT PROGRAM

In the 1965 Medical Library Assistance Act, Congress proposed to foster research and investigations in medical library science and related fields in the interest of improving biomedical information services. With appropriations of $6 million for 1965 to 1970, NLM funded 103 projects concerning the development and evaluation of information activities in libraries, studies in the broad field of biomedical communication, and historical studies of matters related to health and medicine. Because most medical librarians were inexperienced in research, and few of the projects led to applying and implementing new modes of biomedical communication, in 1970 the Director of NLM concluded that such expenditures were among the least rewarding in the extramural program (37).

In succeeding reauthorizations, Congress emphasized that its research interest was in advancing the science of health communications. In 1970, it added an amendment permitting support for demonstration projects for new techniques, devices, or systems that were ready for application, and later added the authority to support projects for the development of new techniques and materials for processing and disseminating health information (Public Law 91-212).

Clinical librarianship, a successful and well-known project funded by the program, was initiated by Gertrude Lamb at the University of Missouri in Kansas City in 1972 (6). The clinical librarian provides information services in a patient care setting as part of a patient care team. As part of the team, the librarian is intimately acquainted with the health professionals' information needs. Although the specific functions performed vary with the medical or surgical service and with the medical facility, the basic design is generally similar: medical librarians accompany physicians on daily rounds and attend weekly staff conferences, noting particularly difficult aspects of individual cases.

The librarian then conducts a literature search, using manual and on-line methods, including MEDLARS, selects a few relevant articles, and provides them to the attending physician(s). Clinical librarians often teach courses in information techniques to both students and teachers in medical facilities. Since the initial research and development grant was awarded, the clinical librarian program has been incorporated into 120 medical schools and teaching hospitals (75).

In a similar vein, the Cleveland Health Sciences Library at Case Western Reserve University began the Circuit Librarian Program in 1973, linking suburban hospitals to its resources. On a regular (usually weekly) basis, medical librarians visit hospitals in surrounding communities, taking information requests from physicians and nurses, ancillary departments, and administrators. Before returning to a hospital, a librarian will have spent time at a resource library filling requests for biomedical and health care literature and audiovisual items. Circuit librarians also assist hospitals in developing in-house collections of core medical literature. The program has been adopted by a number of resource and large hospital libraries, often with NLM grants (10). Many programs, including that of the Cleveland Library, are now self-supporting, with costs covered by the hospitals receiving service.

Although generally considered successful, the research grant program did not achieve all its goals. In 1978, an NLM task force evaluated the program, recognized its many contributions to the biomedical communications process, and applauded the quality of the projects the program supported (131). Nevertheless, the task force was concerned that too few first-rate grant applications were being submitted to the Library to assure adequate advances in the state of the art; that long-range commitments to improve biomedical communications in research and training were lacking; and that potential applicants, staff, and consultants were uncertain about program goals and objectives.

Partially as a result of the recommendations made by the task force, NLM made a number of changes in its research grant program. It now uses four grant mechanisms, similar to those used throughout NIH, to fund research activities:

1. Program project grants—clusters of research efforts having a common focus with coordination by a senior principal investigator.
2. Research project grants—single projects initiated and directed by a single investigator.
3. New investigator research grants—small awards for the young investigator with less than 5 years experience since obtaining a doctorate.
4. Research career development awards—awards providing salary and related support for promising researchers to devote full-time to research for 5 years.

The last two categories were added to augment the supply of research manpower capable of advancing biomedical communications, and were recommended by the task force.

The Library has identified three areas of interest for these grants: new methods for the representation of medical knowledge; classifying, indexing, and abstracting information; and user needs and behavior.

The task force had indicated the urgent need for research in these areas. However, applicants can propose to conduct research in all the areas identified in the Medical Library Assistance Act, which include medical library science, computer technology, biomedical communications, and the history of medicine and related health sciences.

Another innovation, prompted by the 1978 task force report, is the Computers in Medicine Program, a subset of the total NLM research effort. It emphasizes computer science research in knowledge representation, data base management, and clinical decisionmaking (108). In 1980, Congress designated $1.3 million specifically for this program. The Library uses the same grant mechanisms in this program as in the entire research grant program. All the current new investigator grants, the research career development awards, and 9 of the 27 research project grants are in this program.

Table A-7 illustrates the funds distributed for the research grant program and the number of projects funded. For 1980 and 1981, the funds stabilized at about $2.7 million per year and the number of projects remained constant at about 30. In fiscal year 1982, the Library expects to fund 27 projects with $2.3 million.

SPECIAL SCIENTIFIC PROJECTS

As conceived in the original 1965 legislation, the special scientific grant program was to assist established researchers in reviewing, evaluating, and synthesizing extensive collections of medical literature. The recipient was expected to devote full-time to examining the scientific record in a field relevant to the established programs, and to produce a thorough, book-length literature review. Although the special scientific grant program has varied little in its 15-year existence, the 1970 legislation changed the funding mechanism from fellowships to grants in recognition of the program's research orientation, and allowed awards to institutions as well as individuals (Public Law 91-212).

The current program supports qualified scientists and practitioners preparing comprehensive analytical and interpretive documents on major health topics. Investigators are expected to bring together dispersed literature in a subject area, or bridge different subject areas, in the health field. In this way, other health professionals obtain easier access to the continuously expanding biomedical literature. In almost all cases, the work produced has a limited audience and as such would not be of interest to a commercial publishing house. Indeed, one of the criteria used in awarding a grant is that the proposed publication be commercially nonviable.

The program was and remains small in respect to the total funds expended in the extramural programs and the number of projects supported. In 1965, it was expected that the $2.5 million authorization would support approximately 125 medical scholars for the 5-year period. Instead, appropriations totaled $200,000 and only 10 fellowships were awarded. Although the funds appropriated and the number of awards more than doubled in the next 10 years, they still represented only 41 grants totaling $1.2 million (see table A-8). There are seven grants totaling $290,000 scheduled for 1981 and two grants estimated at $23,000 for 1982.

Table A-7.—Medical Library Assistance Act: Research Grants,[a] Fiscal Years 1971-82 (dollars in thousands)

Year	Amount	Number of awards
1971	$ 590	19
1972	640	24
1973	608	26
1974	875	22
1975	1,292	20
1976	1,353	17
1977	1,180	15
1978	1,111	13
1979	1,593	21
1980	2,724	31
Subtotal	11,966	208
1981	2,774	31
1982	2,257	27
Total	16,997	266

[a] Includes some publication grants awarded with research grants, fiscal year 1971-75.

SOURCE: National Library of Medicine.

Table A-8.—Medical Library Assistance Act: Special Scientific Project Grants, Fiscal Years 1971-82 (dollars in thousands)

Year	Amount	Number of awards
1971	$ 5	1
1972	100	4
1973	76	3
1974[a]	95	3
1975	153	4
1976[b]	72	3
1977	109	3
1978	248	7
1979	214	7
1980	142	6
Subtotal	1,214	41
1981	290	7
1982	23	2
Total	1,527	50

[a] Includes release of fiscal year 1973 impounded funds.
[b] Includes transition quarter funds, July 1-Sept. 30, 1976.

SOURCE: National Library of Medicine.

Despite the program's moderate funding, the reports accompanying the 1974 and 1978 legislation considered it successful in enabling senior health professionals to analyze and synthesize biomedical literature, and produce and disseminate biomedical publications of a nonprofit nature.[8] [9]

There are currently 13 active special scientific projects ranging in size from $2,160 to $82,529, and covering a variety of health subjects, including disclosure and consent in medical and legal practice; control of infectious disease in the 20th century; and environmental hazards to small children.

PUBLICATION GRANTS

Publication grants support the preparation and publication of scientifically significant secondary manuscripts—such as indexes, critical reviews, and monographs—to aid health professionals in obtaining relevant literature. The grants are limited and short term, and support projects that NLM believes are important, but whose products attract only a few select readers. These scientific publications are not commercially viable, and have no alternative source of support.

With the uninterrupted growth of published primary biomedical literature, the need for such secondary literature is as pressing today as when the Medical Library Assistance Act was first passed in 1965. At that time, legislators saw a need for supporting publications other than original articles. Interest in this area remained through the five reauthorizations. Although the appropriations were always considerably less than the authorized funding, the report accompanying the 1970 reauthorization congratulated the Library for its efforts and accomplishments in funding the development and publication of over 150 bibliographies, critical reviews, handbooks, translations, and other monographs in biomedical communications.[10]

The original legislation clearly differentiated publication grants from special scientific project grants, in that the former were to be awarded only to medical or scientific scholars to synthesize a body of literature related to their particular research topic, and who wished to devote full-time to this enterprise. Publication grants focused on the publication of biomedical information in forms other than journal articles. The lines between the two programs appear to have blurred, and today the emphasis of both authorities is on providing grants for critical reviews. The staff of the Extramural Program Division have a favorable attitude toward this merging, as they consider the production of critical reviews "one of the major purposes of the extramural programs." Such publications synthesize, and thus provide access to, billions of dollars of biomedical research findings, much of it funded by the Federal Government, primarily through the National Institutes of Health.

The funds appropriated under this authority and the number of grants awarded are displayed in table A-9. Funding increased from $280,000 in 1971 to a maximum of over $1 million in 1978. Since then it has decreased, with a 1982 budget of $514,000. Currently, there are 49 active grants with awards ranging from $500 to $111,839.

Regional Medical Library Program

The mission of RMLP is to provide health science practitioners, researchers, educators, and administrators with timely, convenient access to health care and biomedical information resources, through a coordinated network of health science libraries and information centers. Specifically, RMLP's objective is "to assist in the development of a national system of regional medical libraries, each of which would have facilities of sufficient depth and scope to support the services of the medical libraries in the region served by it" (Public Law 89-241). Although RMLP is now organizationally separated from other Medical Library Assistance Act programs, it shares the general objective of improving information services in the health field, and serves as the focus of many Medical Library Assistance Act program activities.

Table A-9.—Medical Library Assistance Act: Publication Grants, Fiscal Years 1971-82 (dollars in thousands)

Year	Amount	Number of awards
1971	$ 280	16
1972	311	19
1973	389	20
1974[a]	451	25
1975	614	36
1976[b]	668[c]	44
1977	773[c]	43
1978	1,069[c]	47
1979	795	36
1980	787	35
Subtotal	6,137	321
1981	818	34
1982	514	19
Total	7,469	374

[a] Includes release of fiscal year 1973 impounded funds.
[b] Includes transition quarter funds, July 1-Sept. 30, 1976.
[c] Includes Council of Library Resources Training Contract.

SOURCE: National Library of Medicine.

---

[8] U.S. Congress, Senate Report 93-764, op. cit.

[9] U.S. Congress, Senate Report 95-838 accompanying S. 2450, Human Resources Committee, May 15, 1978.

[10] U.S. Congress, Senate Report 91-480, op. cit.

When RMLP was originally conceived, the NLM Director recognized a need to decentralize some of the Library's activities into regional arrangements. The inadequate state of medical libraries and the problems in communicating up-to-date information to practitioners and researchers were described in hearings on the need for the Medical Library Assistance Act. Through RMLP, NLM sought (and continues) to encourage strong libraries to share collections and expand services, so that resources will be available *locally* to meet local and regional needs. The program is intended to increase access to the biomedical literature, particularly for health professionals remote from libraries of excellence, through improving immediate resources and developing a network of backup resources, while avoiding duplicating extensive or specialized collections which are not needed as local resources.

NLM implemented the program by awarding grants to libraries with existing resources and regional services that could be expanded without interrupting normal local services. Such libraries serve as a link between NLM and local libraries. As may be expected, most libraries initially designated as Regional Medical Libraries were in academic institutions. Each was to provide seven basic services to the libraries in its region, the most important being the free loan of books and photocopies of journal articles (i.e., document delivery). Other services include MEDLARS searches, traditional or manual reference services, evaluation of regional information needs and resources, training and orientation, publicity about RMLP, and continuing education for health professionals about sources of information.

Document delivery was initially emphasized, especially for the journal literature, because it was the most effective way to meet the program's mission. The specifics of program implementation, however, differ from region to region to match the needs of users and the characteristics of the regional library. Document delivery (or interlibrary loans) remains a fundamental service of the program.

The first grant was awarded in 1967 to the Countway Library at Harvard University, and by 1970 all 10 planned programs had been started. (The 11th regional medical library is NLM itself, which serves the mid-Atlantic region.) In 1970, Congress noted that progress toward developing a large, coordinated, and cooperative program had been "encouraging; the response of the health library community in its efforts to work with this program [had] been enthusiastic."[11] By succeeding reauthorizations, Congress has reasserted its appreciation of the program's achievements.

RMLP emerged as a four-tier pyramid, with each tier serving as a backup resource for the one below, particularly for document delivery. At the peak of the pyramid, NLM provides policy and planning at the national level in addition to its backup function. At the next level, the regional libraries implement the national policy, coordinate regional library services and educational activities, publicize the program, and provide backup for document delivery. Each regional library has considerable responsibility and freedom of action in the management of the program in its region. As a result, the program has developed differently in each region to match both the needs of users and the characteristics of the regional library. In all regions, however, the interlibrary loan program has a high priority.

The next tier is that of the resource libraries, which subcontract from the regional medical libraries for some of the services they provide. Currently, there are about 100 resource libraries, located mainly in medical schools. Their major function is to fill interlibrary loan requests from the basic unit libraries, the fourth tier of the pyramid. When one resource library cannot fill a request, it transmits the request to another resource library in its region or to the Regional Medical Library. In addition, resource libraries assist with the coordination of network development and with educational activities.

The basic unit libraries at the base of the pyramid are mainly hospital libraries, although any health-related library is eligible to become part of RMLP. Approximately 3,800 basic unit libraries now participate. Their financial responsibilities were originally limited to only those costs associated with communicating with the resource libraries, but they now pay part of the costs of interlibrary loans as well. Basic unit libraries are an essential element in RMLP because they are usually the entry point to the network for the health professional. In fact, the relationship between RMLP and the other extramural programs authorized by the Medical Library Assistance Act is most evident at this level. As noted earlier, many resource improvement grants have been used to encourage community hospitals to develop collections of text books and journals, share resources, and become active participants in RMLP.

The vertical, pyramid structure of RMLP is organizationally important for the sharing of core biomedical information resources. But as the health sciences' knowledge base broadens, and health care delivery encompasses evermore diverse and specialized subjects from the social sciences, law, and economics, the *horizontal* sharing of resources is becoming increasingly important, especially for the level of sophistication found in the resource and basic unit libraries.

---

[11] U.S. Congress, Senate Report 91-480, op. cit.

Rather than looking to libraries above them in the RMLP pyramid, smaller libraries are beginning to tap specialized resources *outside* the pyramid for materials outside the traditional field of biomedicine and unavailable in medical libraries. While not a program objective, this horizontal sharing of resources has been encouraged by the existence of RMLP (27).

Although RMLP has retained the same general objective since its inception in 1965, program format and activities have been modified in response to legislative and administrative demands, user needs, level of each region's network development, and technological change. The 1970 extension of the Medical Library Assistance Act permitted the use of contracts as well as grants for financing the program, and by 1972, all awards to regional medical libraries had been converted from grants to contracts. The contract mechanism was again modified in 1979 according to DHHS regulations to require competitive bidding among regional medical library applicants. Eight institutions were awarded contracts in fiscal year 1980 under this new method; a ninth contract was awarded the following year. The only site change resulting from this new process was in the Midwest region, where the University of Illinois replaced the John Crerar Library as the Regional Medical Library.

In 1972, NLM issued a policy statement that committed the Library to the development of a Biomedical Communications Network (BCN) and described RMLP as the first phase (106). The long-term objective of RMLP was to serve as a model for BCN, which would be designed for information transfer supporting health services delivery, education, and research. Its immediate objective was to develop a document delivery system for the Nation's medical libraries. In 1979, a Library committee reaffirmed RMLP's general objective, but broadened its immediate objective to include encouraging greater resource sharing and providing services beyond document delivery.

A cost-sharing plan for document delivery was also implemented in the RMLP in 1979, based on NLM's belief that local libraries should bear the financial responsibility for documents provided to their primary users. The concepts of the plan were developed in 1970 when the Library determined "that it [was] not wise to base the finding of the entire [Regional Medical Library] network upon appropriated funds." Cost sharing will gradually eliminate NLM's financial commitment to document delivery by October 1982, leaving the Library responsible only for those materials not available in the Regional Medical Libraries and allowing it to support other aspects of resource sharing and the development of better communication links.

Cost sharing has already permitted NLM to decrease its funding level for document delivery from 50 to 25 to 30 percent. Because of the increasing volume of loans along with the rate of inflation and the desirability of having material as close as possible to the requestor, the Library concluded that it was essential that local libraries exercise greater fiscal responsibility.

The notion of cost sharing for document delivery services has been controversial since it was first raised in 1970. Some regional libraries now view the plan as a shift on the part of NLM away from funding interlibrary loans and towards funding the development of MEDLARS III, the latest version of the Library's online retrieval system. Because MEDLARS III is primarily designed to alleviate the Library's internal burden of managing a growing body of medical literature, these regional libraries see the Library withdrawing support from the Regional Medical Library network at a time when new demands are levied against their resources by small libraries with newly acquired access to MEDLARS.

NLM insists that its reduction in direct financial support for document delivery does not equate with a "deemphasis" of that activity. Further, the Library notes that an important feature of MEDLARS III will be the full automation of interlibrary loan referrals, a development that will eventually enhance document delivery in that it will be easier to identify and locate a bibliographic entity. Further, development costs for MEDLARS III are drawn from the Library's operating budget, not from funds authorized for the Medical Library Assistance Act.

In October 1981, NLM announced a reconfiguration of the Regional Medical Library network, reducing the 11 geographic regions to seven effective November 1982. The proposed change is intended to reduce the administrative costs of the program to make more funds available for program activities, and is in response to congressional sentiments evident in 1981 hearings on the reauthorization of the Medical Library Assistance Act. As library and information services have become increasingly computerized, larger geographic areas have become easier to manage, allowing NLM to consider redrawing boundaries as a more cost-effective mechanism to meet current and anticipated budget constraints (94). Along with the new regions, the Library proposes to establish a national network advisory board, similar to those presently in place in each region. NLM believes this commission will allow users of RMLP's resources and services to be more involved in the program.

Appendix B.—U.S. Information Policies

This appendix briefly reviews the history, current standing, and future prospects of domestic information policy issues to provide a larger context for issues considered in the report. The relationship between domestic information policy and international information policy is noted as well.

The United States does not have a uniform, articulated, national information policy guiding the information activities of the country and, so for the most part, these activities have evolved informally. However, the Government has long been involved in creating, gathering, organizing, and disseminating information.

Three separate constitutional provisions define the role of the Government in relation to publishing. The first amendment denies to the Government the right to abridge freedom of publishing. The other provisions give the Government a specific role in the *promotion* of publishing. The copyright clause empowers Congress to give authors exclusive rights to their products, and the postal clause allows Congress to establish post offices and post roads.

From colonial days through the 19th century the Government supported development of the press in various ways. Postmasters were often the first newspaper publishers, because they had a source of news and a means of distribution. Benjamin Franklin, the first Postmaster General, was the most famous of the postmaster-publishers. Official notice advertising and Government printing was used to support newspapers. Thomas Jefferson, finding no Washington newspaper supporting his administration, brought the Philadelphia publisher, Samuel Harrison Smith, to Washington to start the *National Intelligencer,* and had Congress give it printing contracts. Congress gave newspapers free local distribution through the post office until it substituted second-class mailing rates in the mid-19th century, a revised form of subsidization.

Despite this early concern with information, it was not until after World War II, with the extraordinary growth in science and technology, that the United States expressed a sustained interest in developing a coordinated information policy. As a result of the increased tempo of scientific activities, scientific knowledge had expanded and become unwieldly, and scientists had increasing difficulty keeping informed of new developments. In searching for a way to facilitate full and open communication among scientists, the Federal Government sponsored a series of reports on information policy issues. A common thread running through the studies from 1958 to the late 1960's is the need for improved and coordinated science information services, particularly on the part of Federal agencies. The studies devoted little attention to the private sector, and then only to individuals and the nonprofit professional societies.

By the mid-1960's the Federal Government had begun to view information as a national resource. Although it remained interested in the improvement and coordination of scientific and technical communication, the Government gradually broadened its perspective on information activities. Information in many fields, such as commerce and law, and varied aspects of information, information communications, information technology, and information economics, were incorporated into the national dialog. At the same time, the private sector began to assume more importance in the policy debate, and the focus shifted from the nonprofit to the profit component of the private sector.

Today, information is so central in our social order, that some observers have termed the United States a "postindustrial society," a "knowledge-based society," and an "information society." Accordingly, information policy concerns have increased rapidly in number, diversity and complexity.

In 1976, information policy was described as having many connotations, including "policy dealing with the regulation of information messages over common carrier facilities, policy with respect to postal rates for the distribution of books throughout the country, policy affecting the information requirements imposed by Federal and State governments, and policy concerning the communication of research results to the scientific and technical community in the public and private sector" (144). By 1981, an OTA report focusing exclusively on computer-based national information systems noted (115):

> It would not be possible for any one study to capture succinctly a single set of policy issues that would apply to all national information systems in American society. The specific system applications are too diverse, the potentials and problems too complex, and the parties-at-interest and relevant institutions and legal frameworks too diverse.

The OTA report (115) did identify current policy issue areas concerning national information computerized systems, including:

- privacy;
- security;
- Government management of data processing;
- society's dependence on information systems;
- transborder data flow;
- information gap;
- innovation, productivity and employment;
- constitutional rights;

- computer crime; and
- regulatory boundaries; and computer software protection.

Many of these issues are not confined to computerized information systems, but extend to other information activities. Some pertain to information activities in the international arena.

The OTA report (115) also noted the absence of a uniform, coherent national information policy. The type of information policy or policies, domestic and international, that would best serve the interests of this country has received considerable attention, but still remains elusive. Most studies regard a national information policy, as prerequisites for effective information pursuits, although it is not a universally held position.

If a national policy were developed, many structural questions would still be undecided, among them: should there be a uniform, monolithic policy for all information activities; should there be a diverse set of policies to account for the wide variation in information activities and issues; should the power for designing and administering domestic or international information policy be centralized or decentralized; if centralized, where should its locus be; and, if decentralized, which Government agencies and departments should be involved.

The present pluralistic and undefined domestic information policy is considered by some to hinder the ability of the United States to devise and effect strategies concerning information and communication in the international arena. Whereas the majority of foreign administrations dominate their domestic communications activities, in the United States, communication and information functions are exercised by both the Government and the private sectors. Although this pluralistic approach to communications and information has contributed to the eminent position of the United States in telecommunications, it has complicated the definition of a domestic and international information policy. In addition, international information policy is tightly intertwined with foreign policy, and is viewed as a negotiating tool from some perspectives.

The formulation of an international information policy is becoming more and more important as technological developments stimulate international information activities. Projections suggest that within two decades an integrated global network for information transmission will be developed and that a person with a universal terminal will be able to access the network from almost any place in the world to obtain information on nearly any subject (127). Other technologies, such as video disks, may increase the ease of transmitting information globally even earlier, with less need for international telecommunication networks (133). These developments force attention to many international information issues, such as privacy, copyright and restraint of trade.

The current responsibility for information activities is fragmented in numerous Government agencies. There is no coordinating mechanism for intragovernmental policy issues, nor is there an opportunity or locus for Government and private dialog about matters of mutual concern. The Paperwork Reduction Act of 1980 (Public Law 96-511) gives the Office of Management and Budget (OMB) the major responsibility for regulating executive agencies regarding information acquisition and distribution; OMB directives concerning information activities appear to be focused on saving the Government costs in a time of extreme budgetary constraints and placing reliance on the private sector for goods and services needed by the Government and U.S. citizens. They do not provide the broad perspective required to address the complex issues of an information age.

Because of the interlocking nature of these international and domestic information issues, it has been suggested that they be addressed in a coordinated fashion by a high-level centralized executive body. However, the prospect for an integrated approach to information issues appears small. Rep. George E. Brown, Jr., notes the need for involving the executive and congressional branches in an understanding of information policy issues (23):

> Few members of Congress appreciate the potential contributions and consequences of information science and technology. Nor are they aware . . . of policy issues and the need to plan for the changes to come.

A 1981 OTA study (115) supports this conclusion and comments on the lack of interest among present policymakers in a uniform Federal information policy to address the many problems that might conceivably occur from the use of data systems.

List of Selected Domestic Information Policy Activities

1952.—Title V of the Independent Office Appropriations Act (31 U.S.C. 483a) provides that agencies set prices to recover as fully as possible the entire costs of providing a service, taking into account the public good and the benefit to the user. Although the Independent Office Appropriations Act is not confined to information services, charges for data bases and dissemination services fall under its purview. The guidance it provides is very general and open to interpretation.

1956.—The National Library of Medicine (NLM) is established as a national library and resource for biomedical information by the National Library of Medicine Act (Public Law 84-941).

1958.—The Baker Report (122) affirms the need for a free flow of scientific information and recommends a Federal research and development coordination mechanism.

1959.—OMB issues Circular A-25 (48) to implement the Independent Office Appropriations Act by requiring that charges be made to each identifiable recipient of a Government service from which a special benefit is derived, and that the Government recover full costs in rendering the service. Exceptions are made when the activity is designed for public safety, health or welfare. The circular is difficult to interpret with respect to who and how much should be charged.

1960.—The Subcommittee on Government Reorganization and International Organization's report, "Documentation, Indexing and Retrieval of Scientific Information" (147), recommends improved Federal and private information services, and strengthened coordination.

1962.—The Crawford Report (142), prepared for the Federal Council on Science and Technology, recommends that each agency have a single office responsible for science information activities.

1962.—The Surgeon General's Conference on Health Communications issues a report (138) on the need for improved communications of scientific research results, research training, and the use of libraries as communication centers and resources.

1963.—The Weinberg Report (120) asserts that the Federal Government should assure the ready availability of information concerning research in progress, through a network of Government information centers of the Federal Council for Science and Technology. Nongovernmental information systems were to be examined for overlap with Government systems.

1963.—The National Academy of Sciences/National Research Council's report *Communications Problems in Biomedical Research* (111) stresses the biomedical community's responsibility for improved facilities, research and development, training, and coordination for the biomedical information complex.

1964.—The Committee on Scientific and Technical Information (COSATI), established by the Federal Council on Science and Technology to coordinate a wide range of activities, commissions a report (145) examining alternative means for developing a Federal information program. Because of its charter, COSATI focused totally on the governmental components.

1965.—The President's Commission on Heart Disease, Cancer, and Stroke stresses the need for better communication of biomedical research and improved medical libraries to prevent the loss of new scientific knowledge (121).

1965.—The Medical Library Assistance Act (Public Law 89-241) is passed to improve the production and dissemination of information in the health field.

1967.—A report by Stafford L. Warren (158) to the President on a National Library of Science System recommends that NLM be one subsystem of an overall science system.

1968.—The Lister Hill National Center for Biomedical Communications is established as part of NLM to improve biomedical communications through advanced technologies.

1969.—The National Academy of Sciences and the National Academy of Engineering jointly commission the Scientific and Technical Communication (SATCOM) Report (95). The report calls for a nongovernmental body to be responsible for national information policy, at least in the area of science and technology, as a joint commission of the National Academy of Sciences and the National Academy of Engineering. The SATCOM Report explicitly recommends close cooperation between the public and private sectors in future development, and suggests Government financial support for information services operated by professional and scientific societies. The needs of the for-profit sector are less clearly defined. Coordinating mechanisms for the Government (COSATI) and the private sector (SATCOM) are established to promote interrelated activities.

1971.—The Kozmetsky Report (73), explicitly recognizes the general value of information as a critical resource, both nationally and internationally, rather than limiting it merely to the realm of scientific and technical contexts.

1972.—The Greenberger Report (59), done for the Federal Council on Science and Technology and the National Science Foundation, examines the role of COSATI, and concludes that the Government is not well organized to deal with the problems in developing information as a national resource. It recommends that new policy mechanisms be created and that the private sector have input in policy and program development.

1975.—The National Commission on Libraries and Information Science (NCLIS) presents *Toward a National Program for Library and Information Services* (97), a report based on extensive public hearings and meetings throughout the country. This report considers a broad range of information needs, including those of the general public, science, technology, business and industry, and education. Like earlier reports, it identifies the need for cooperation among the several sectors in developing information as a national resource. It also suggests a governmental role in providing technical inducements and finding incentives for the private sector and State governments.

1976.—The Oettinger Report (114) reviews issues related to fragmentation, confusion, and contradiction in present Federal information policies.

1976.—The Becker Report (15), done for the National Science Foundation, summarizes the history of information policy and reiterates the need for a coordinated national effort, with Government sponsorship for an institute for information policy and program planning in which private organizations can be accommodated.

1976.—The report *National Information Policy* (144), prepared by the Domestic Council Committee on the Right of Privacy, is published by NCLIS. It identifies issues affecting the relationship between the Government and the private sector in the production and dissemination of information and argues for a standard set of policies clarifying the relationship between the two sectors. It recommends a strong information policy group in the Executive Office of the President and the creation of appropriate intergovernmental and non-Federal committees.

1976.—The SCATT Report (2) provides a comprehensive plan for integrating the various public and private components in the production and dissemination of scientific and technical information.

1978.—*Into the Information Age: A Perspective on Federal Action on Information* (60), a study commissioned by the American Library Association, calls for Federal leadership in developing information to meet societal needs such as air quality, energy, economic well-being, public safety, and environmental preservation.

1978.—The National Telecommunications and Information Administration is established in the Department of Commerce to examine broad information policy questions.

1979.—OMB distributes Circular A-76 revised (49), "Policies for Acquiring Commercial or Industrial Products and Services Needed by the Government." It declares that the general policy of the Government is to "rely on competitive private enterprise to supply the products and services it needs."

1979.—The Library of Congress surveys the publications policies of executive branch agencies.

1979.—The congressional Joint Committee on Printing publishes *Federal Government Printing and Publishing: Policy Issues* (143), identifying issues with respect to: 1) administration of policy, 2) Federal Government printing production and procurement, 3) impact of new technology, 4) access to and distribution of Government information, 5) the depository library program, and 6) the pricing of Government information.

1979.—The General Accounting Office publishes, *Better Information Management Policies Needed: A Study of Scientific and Technical Bibliographic Services* (55), confirming the need for better Government management of information centers, and identifying duplicative services and facilities, failures to recover costs, and inconsistent cost recovery procedures among agencies. It recommends that the Director of OMB direct each department and agency to designate a high level official responsible for information management, that Congress consider more precise language when authorizing information centers to alleviate duplication and provide more specific guidance on which information services should be exempt from cost recovery requirements.

1979.—The President announces measures to "help ensure our country's continued role as the world leader in industrial innovation" including "enhancing the transfer of knowledge" and "increasing technical knowledge." One of the actions taken to ease and encourage the flow of technical knowledge and information establishes the Center for Utilization of Federal Technology at the National Technical Information Service (Department of Commerce) to improve the transfer of knowledge from Federal laboratories and, through the Departments of State and Commerce, to increase the availability of technical information developed in foreign countries.

1980.—OMB Bulletin No. 81-16, "Elimination of Wasteful Spending on Government Periodicals, Pamphlets, and Audiovisual Products," imposes an immediate moratorium and institutes a comprehensive review of the production, procurement, and dissemination of new audiovisual products, periodicals, and pamphlets, and calls for user fees to recover the costs of production.

1980.—An OMB draft circular, "Information Management and Dissemination of Federal Information," outlines a cost recovery program for information provided by the Government. The proposal covers all costs associated with dissemination, including printing, processing, and retention, but excludes the cost of producing or creating the information. The policy is not made official, though some aspects are included in the Paperwork Reduction Act of 1980.

1980.—The Working Group on Private Sector/Government Relationships for Scientific and Technical Information, of the Federal Coordinating Committee on Science, Engineering, and Technology identifies the four key issues in private sector/Government relationships: 1) the different philosophical views of information as a resource; 2) determination of Government's legitimate role in the operation of services for a given philosophical view; 3) the historical and future role of Government as a risk-taker in the development of technologies and markets; and 4) the kind of platform or mechanism through which agencies and the private sector can resolve differences. The group concludes that attempts to develop guidelines for resolving differences across Government agencies and between sectors is not feasible, as neither sector acts and reacts as a single entity.

1980.—OMB issues Circular A-121 to "establish policies that promote effective and efficient management and use of certain data processing facilities" by instituting business-like procedures of cost accounting, cost recovery, and interagency sharing of data processing facilities.

1980.—The Paperwork Reduction Act (Public Law 96-511) establishes an Office of Information and Regulatory Affairs within OMB to regulate and coordinate the activities of executive branch agencies with respect to information acquisition and distribution. It requires that each agency have a single authority, at the assistant secretary level, responsible for information activities within the agency. It also authorizes the establishment of a Federal Information Locator System, to be composed of a directory of information resources, a data element dictionary, and an information referral service.

1981.—*Issues in Information Policy* (38), a report of the National Telecommunications and Information Administration (Department of Commerce), divides information policy issues into two groups: 1) those encompassing constitutional and statutory authority for permitting, requiring, or inhibiting the availability and accessibility of information; and 2) those focusing on economic policies for inhibiting, managing, or facilitating the distribution of information to certain sectors of society.

1981.—The OTA report, *Computer-Based National Information Systems: Technical and Public Policy Issues* (115) describes how future applications of computerized information systems may intensify or alter the character of the policy debate and the need for new or revised laws and policies.

1981.—The Public Sector/Private Sector Task Force of NCLIS (see app. F) concludes that the four major issues in the conflict between the sectors rest on the need for the Federal Government: 1) to take a position of leadership in facilitating the development and fostering the use of information products and services; 2) to encourage private sector investment in information resources, products and services; 3) not to engage in commercial information activities unless there are compelling reasons for it to do so (and there must be well-defined procedures for determining that such reasons indeed are present); and 4) to protect private sector property rights in any package of governmentally distributed information, that includes private information resources, products, or services.

1981.—OMB releases Memorandum 81-14, providing criteria for the evaluation of Federal information centers by executive departments and agencies conducted under the Paperwork Reduction Act. The criteria are used to determine whether an information center duplicates private endeavors or can be consolidated with other centers, and whether centers should provide information on a full cost recovery basis.

Appendix C.—MEDLARS Evaluations: A Review of the Literature

Introduction

This appendix reviews published and unpublished studies evaluating MEDLARS services and usage. In many ways, these studies have been dated by a decade of rapid technological advances in computerized retrieval systems. A paucity of studies is available on the system in place today; thus, generalizations based on the findings presented here may be misleading. The literature review here is offered to illustrate the kinds of questions one must ask in evaluating on-line retrieval systems, and the complexity and difficulty inherent in such an undertaking.

The National Library of Medicine (NLM or the Library) has devoted extensive time and other resources to internal and external review and has examined not only system performance but the basic purposes, goals, and constituency groups MEDLARS is designed to serve. Few institutions, public or private, have matched NLM's commitment to evaluation. This appendix reviews studies of user satisfaction, searcher variation, and the retrieval capabilities of the MEDLARS system. It also presents OTA's evaluation of MEDLINE's coverage of selected topics. First, however, it describes the types, number, and motivations of MEDLARS users.

Users and Usage of MEDLARS

NLM collects utilization data for MEDLARS solely on an institutional basis. There is information on how many medical schools and hospitals have MEDLARS terminals, but there is relatively little information on the individuals requesting on-line searches, their reasons for seeking the information, or their level of satisfaction with search results. Further, and perhaps more importantly, there is no study available of the nonusers of MEDLARS and their reasons for not using its services.

Because so few data are available, users' expectations can only be appraised in light of their organizational affiliations, and the respective orientation of those institutions. In September 1980, 1,243 institutions in the United States had direct access to NLM's data bases. Forty-one percent of these institutional users were hospitals and clinics, 23 percent were commercial firms, and 9 percent were medical schools (see table 4 in ch. 2 for complete data).

This discussion defines the librarians and information specialists who conduct searches at computer terminals to be *intermediate users* of information systems, while *end-users* are the researchers and clinicians who request that searches be done and who utilize the information retrieved. Sometimes, though rarely, end users conduct their own searches. In 1977, an estimated 80 percent of all searches were conducted by trained intermediate users alone, 20 percent by intermediate users with end users present at the terminal, and less than 1 percent by unassisted end users (160).

Patient Care Institutions

Hospitals are the largest and fastest growing group of institutional users of MEDLARS services. In 1978, hospitals represented one-third of NLM's on-line institutional users; by 1980, this figure had grown to over 40 percent. There are more than 7,000 hospitals in the United States; over 700 have MEDLARS terminals, and many others obtain on-line services from hospitals in nearby communities. In 1976-77, one study indicated that hospitals with terminals each processed an average of 493 MEDLINE searches per year (160). For that same period, professional schools conducted an average of 1,429 searches. The average for all institutions with access to MEDLINE was 842.

Hospital librarians report that about half their search requests are from physicians wanting information directly applicable to patient care. Often, the information is needed to aid the diagnosis and treatment of disease. In such instances, physicians are generally looking for a few relevant articles and need to obtain them within 24 hours. MEDLARS is also helpful to the physician preparing an article or lecture, or simply as a way of the physician keeping abreast of the literature on a particular subject or specialty. Nurses are also frequent users, often for reasons similar to physicians.' Questions on patient management represent the most common queries, though nurses also request searches for patient education, staff development programs, and the preparation of papers and presentations. MEDLARS is increasingly being used by ancillary service departments and administrative staffs in hospitals, especially for the planning of new services and the purchase of new equipment.

Though MEDLINE is the most frequently used data base in hospitals, the HEALTH and CANCERLIT files are also popular. Twenty-two percent of hospitals with on-line access to NLM's data bases also reported having used other non-NLM data bases in 1977, with PSYCHOLOGICAL ABSTRACTS and SCISEARCH the most often used. As the holdings of most hospital libraries are generally not extensive (the 1977 average number of serial subscriptions was under 400 for hospitals, but over 2,000 for professional schools) and primarily clinically oriented, some document delivery problems are reported for articles identified through non-MEDLINE data bases. Sixty-one percent of all institutional MEDLARS users said they needed to acquire additional holdings based on the demand from on-line searches but did not have the funds to do so (26).

Research Institutions

The majority of institutional users are interested in MEDLARS as a research tool. These users include Government agencies, research foundations, professional associations, medical and professional schools, commercial research and development firms, and information brokerages. Little information is published on the utilization of MEDLARS services by such organizations, though some basic descriptive data has been reported by medical school libraries on their individual users (58,88,139,140,141). Although medical schools constitute only a small portion of research institution users, their experience with MEDLARS reflects that of similar institutions and is particularly useful for identifying the reasons that researchers request on-line searches.

MEDLARS usage in medical schools is characterized by utilization patterns similar to those found in hospitals. Both hospital and medical school libraries conduct searches for physicians, nurses, lab technicians, and administrators, and often for the same reasons—patient care, preparation of articles and lectures, and planning. Medical school libraries, however, see the majority of searches requested for research purposes (58,88,139,140,141).

While health practitioners tend to want a few relevant articles, researchers more often request broad, comprehensive searches. Faculty members, who divide their time between teaching, research, and patient care, tend to be the heaviest users of MEDLARS, requesting information needed for both ongoing and prospective research. For them, MEDLARS is an invaluable means of saving time and effort; for identifying articles that might be missed in a manual search; and for ensuring that they have a comprehensive bibliography available, especially when considering research in a new field (139,141).

Clinical Librarianship Programs

Clinical librarianship programs first appeared in the early 1970's, in response to the need for more timely dissemination of current developments and research findings. Primarily undertaken in large medical centers and teaching hospitals, the programs are designed to bring the skills of medical librarians directly to the delivery of health services by providing "highly specific, case-related medical literature . . . in a manner that permits the information obtained to influence ongoing case management" (14). In a technical sense, the success of such programs depends on the librarian's ability to conduct highly specific, narrowly focused searches, in a brief period of time, ensuring that the relevant information reaches the clinician in time to influence the care of the patient. The ultimate objective of clinical librarianship is improved patient care.

Although these programs have received considerable attention and several articles are available describing their focus, methods, and results (129), evaluative efforts found in the literature have been limited in scope and have not adequately assessed any effects of such programs on patient care. They do report, however, significant increases in utilization of hospital library services resulting from the introduction of clinical librarians. For instance, one study showed a 120-percent increase in search requests from a staff with a clinical librarian program (61). A second study reported that 92 percent of physicians said that they read the articles given to them by clinical librarians, 86 percent said they learned something new from them, and another 20 percent indicated that the articles affected patient management to some degree (129). This study hypothesized that clinical librarianship programs could be a cost-effective alternative to many ancillary services, by substituting relatively inexpensive literature searches for more costly diagnostic tests.

MEDLARS Evaluation Studies

OTA reviewed three types of MEDLARS evaluation studies. The first group of studies, user satisfaction studies, asked end users to determine the relevance of retrieved documents to their requests. End users queried in these studies generally found MEDLARS search helpful in their research and/or clinical practice. A second group of studies evaluated variation in MEDLARS search results based on characteristics of intermediate users. These studies indicate that the training and experience of the searchers and the interaction between intermediate and end users in formulating search requests are key to the success of the on-line search. Finally, Lancaster's landmark study (77) evaluated the retrieval capabilities of MEDLARS before it was available on-line. Lancaster reported average recall and precision ratios of 58 percent and 50 percent respectively, found extraordinarily high variability from question to question searched, and conducted an extensive analysis of recall and precision failure. On the whole, Lancaster found little critically wrong with the system.

The fourth topic discussed below is OTA's evaluation of MEDLINE's coverage of selected topics.

User Satisfaction Surveys

Several librarians have attempted to determine how well MEDLARS satisfied the information needs of end users in their individual libraries. In 1977, Brown (22) reported that users of on-line services, including MEDLARS, were "generally satisfied" with those services. The population studied represented only a fraction of those who could potentially benefit from on-line searches. Underlining the significance of the latter point, 2 years later a MEDLINE feasibility study undertaken at the behest of the Northeastern Consortium for Health Information was based on the theory that "most potential users and supporters of MEDLINE within hospitals [were] unaware of its usefulness and application" (89). A concerted effort to publicize the system's services through the NLM network increased demand significantly. The author concludes that the 13 member hospitals participating in the study could generate sufficient usage to justify the costs of MEDLINE, at least on a shared-service basis.

Sharing access to NLM's data bases was also suggested in a study done on the use of on-line services in academic settings not affiliated with medical schools (11). Here, interest in such systems was identified in students and faculty members engaged in research in disciplines related to the health sciences. In each of these studies, users were given the opportunity to have searches conducted at minimal (or no) cost, and then asked whether they would request searches if charges for on-line time were somewhat higher. The majority of users in each study said that they would pay, although instituting charges was never tested. Nevertheless, by 1977, 90 percent of medical school libraries charged users for MEDLARS services (160).

Several evaluations have been conducted by submitting questionnaires to MEDLARS users, asking them to identify their purposes in requesting on-line searches and their satisfaction with search results. OTA reviewed two efforts undertaken in individual information centers (93,141), and two others that evaluated data combined from seven centers (139,140).

In one study, 246 of 428 users of MEDLINE services at the University of Virginia Medical Library responded to a survey asking whether searches had been of assistance to their research or clinical work, and whether they would continue to utilize the system after the imposition of modest charges. They were also asked if MEDLINE was a "substantial improvement over the traditional methods of searching through the printed indexers" (93). The study group was composed of nursing, medical, and graduate students, nursing and medical faculty, and staff; health professionals outside the medical center were included in the study through a statewide medical information service. Over 93 percent indicated that MEDLINE had assisted in their research and clinical work, and that it was an improvement over manual searching. Seventy-five percent indicated that they would continue using the service after the imposition of charges.

The remainder of the studies reviewed in this section were conducted by Tagliacozzo of the University of Michigan, and are the most sophisticated user satisfaction studies of MEDLARS. In her first study, published in 1973, Tagliacozzo identified two sets of issues that must be considered in evaluating the service provided by any information system such as MEDLARS (141). The first relates to the performance and costs of the system. The second relates to the end users of the system: who they are, whether they represent all categories of users the service was designed for, and whether using the system altered the progress of their research and clnical practice.

The first Michigan study assumed that utilization of an information service could be explained, and to some extent predicted, on the basis of end-users' subjective assessments of its usefulness. Questionnaires were distributed to 275 MEDLARS users. Of 168 completed surveys, 7.2 percent reported their MEDLARS search as "not helpful," 25.9 percent as "moderately helpful," and 66.9 percent as either "helpful" or "very helpful." The respondents listed a variety of reasons for requesting on-line searchers, though clearly the majority of requests was for ongoing or prospective research. Only 24 respondents classified their work as "exclusively clinical," a number far too small to draw any conclusions about the role of MEDLARS as a source of information for practitioners.

In 1975, Tagliacozzo published a second study, specific to MEDLINE, and drawing on data collected in seven Midwestern medical centers (139). The study examined the characteristics, motivations and purposes, expectations, and perceptions of MEDLINE users at a time when the system was just beginning to be used in the medical centers.

Tagliacozzo found that users tended to be either research faculty members working in the basic and clinical sciences, or students working toward advanced degrees or in clinical training. Again, it was difficult to determine the role of MEDLINE in the acquisition of knowledge, though it was apparent that use of the system was not confined to academicians. Most MEDLINE users reported that the retrieved information was primarily for research purposes and that the search was requested because the service was perceived "as a more effective means of reaching relevant citations than the traditional bibliographic instruments."

Tagliacozzo's third study allowed users to distinguish between the value of the search process and the usefulness of the search results, rather than simply offer an overall judgment of the service (140). Data were again collected by questionnaire in seven Midwestern medical centers. Sixty percent of the respondents reported positive reactions to search results, though many others noted that while the search had not provided many useful citations it had saved the time and energy required for a manual search, or had confirmed opinions that all relevant literature for a subject had previously been identified. In her conclusion, Tagliacozzo cautions against taking the users' judgment at face value, especially for determining whether information needs are satisfied through use of an on-line system, because so many factors, such as familiarity with relevant literature and the nature of the search requests, can influence users' responses to the system.

These studies leave the impression that end users find MEDLARS helpful to their work in research and clinical practice. However, for the most part they did not sufficiently test critical questions such as: were the information needs of the user satisfied? and did the system provide the user with all, or most, of the relevant literature which the data base contains?

Evaluations of Intermediate User Variation

Several studies have examined variation in MEDLARS search results based on characteristics of intermediate users. Two studies examined the results of searches conducted by end-users themselves, without the assistance of a trained librarian. One found that "nonlibrarian users" were quite capable of interacting with MEDLINE, as measured by the number of modificatons to their search statements made at the terminal (118). A second study reported usage of two NLM data bases, MEDLINE and TOXLINE, and concluded that, when access to terminals was provided to pathologists and pharmacists, nonmediated use by researchers (i.e., without the assistance of a trained MEDLARS searcher) could be beneficial, if such users were given a "minimanual" describing MeSH vocabulary (130).

A 1978 study compared results of MEDLARS searches conducted through different software packages, one available from NLM, the other from a commercial vendor, Bibliographic Retrieval Services (125). The study reported that searches could be done equally well on either system, but that they differed significantly from a technical standpoint. The differences were reported to be important only to intermediate users and did not affect search results. It was recommended that these variations "justify the dual availability of the files."

A recent evaluation, funded by NLM, examines the effect that the type of user training has on searching style and performance (156). The study considered 535 searches from 191 intermediate users, and found no statistically significant differences in searcher performance between intermediate users trained by NLM and those receiving MEDLARS training "informally." Performance measures for all searches were reported to be 23 percent for recall and 67 percent for precision.

Evaluation of MEDLARS' Retrieval Capabilities

The only intensive effort to determine MEDLARS' ability to retrieve relevant information efficiently was conducted in 1968 before the system was available online. This was a study of batch processing: today's system is very different. Lancaster's landmark study, *Evaluation of MEDLARS Demand Search Service* (77), reported the precision and recall performance of the system for over 300 search requests. Users (practicing physicians and researchers) were asked to assess the relevance of articles retrieved through MEDLARS searches. An article was considered relevant if it had "value to the user in relation to the information need that prompted his request." On the average, Lancaster found MEDLARS to be operating at a 58 percent recall level and 50 percent precision level. He noted that these averages, though characteristic of retrieval systems, might be misleading, since results of individual searches were widely scattered.

Much of Lancaster's study was devoted to an analysis of "search failure:" that is, the reason why searches did not identify more of the relevant literature and why so many irrelevant articles were retrieved. The author reported that 25 percent of the recall failures and 17 percent of the precision failures could be attributed, in part, to a communication breakdown between the user and the system. He recommended that search request forms be redesigned to more accurately reflect the information need of the end user. Changes in indexing, indexing language, and approaches to searching strategy were also recommended.

OTA's Evaluation of MEDLINE's Coverage of Selected Topics

OTA evaluated MEDLINE's coverage of literature relevant to five selected topics in biomedical and health services research. The topics and results are displayed in table C-1. Review articles on each topic were identified, and their bibliographies provided a relevance base of pertinent documents against which MEDLINE's coverage could be measured.

Table C-1.—MEDLINE's Coverage of Selected Topics

Topic	Number of relevant citations	Percent in MEDLINE	Percent recall[a]
Hepatitis and hepatoma.....	161	69%	84%
Hemoglobin genetics......	385	81	94
CT scanners	166	86	82
Patient compliance ...	323	62	77
Health care delivery	423	20	71

[a] Number of items retrieved/number of items included in the data base.

SOURCE: Office of Technology Assessment.

MEDLINE's coverage is more than adequate for the biomedical topics, but less so for those related to health services research. This finding is not surprising, and does not necessarily reflect a deficiency in NLM's system. Documents related and pertinent to the health services topics appear in a wide variety of publications, including law, business and public administration journals, conference reports, and monographs. These publications are not normally indexed for MEDLINE, and it is unreasonable to expect NLM to cover such diverse information sources. Coverage for these topics may be much better in other NLM data bases, especially HEALTH.

Conclusions

Evaluating any information system's ability to satisfy the needs of its users is an enormously complex task, for at the most fundamental level a system's performance is based on human behavioral patterns that do not readily lend themselves to description by simplistic facts and outcome measurements. Human behavior affects every stage of the indexing, search, and retrieval processes—effects that are not reflected in measurements of recall and precision, nor in most studies of user satisfaction. Evaluation procedures specific to information science are being developed, and the study of information needs and systems capabilities is expanding to include communications to other than scientists and researchers, the traditional users of information services. But more creative approaches to evaluation are still needed if information processing and dissemination are to improve (28).

The findings of evaluation studies OTA reviewed are not generalizable to the present MEDLARS systems, nor to its entire user community. Many studies were conducted on earlier versions of MEDLARS, in essence evaluating a system vastly different from that in place today. Others examined a user population or set of search requests too small to carry external validity. Still others paid insufficient attention to the intricacies of the task before them. In the absence of soundly developed evaluation methods applicable to large, complex systems like MEDLARS, the limited value of these studies, though frustrating, is not unexpected.

Appendix D.—Development of Computerized Biomedical Bibliographic Retrieval Systems

In every half-century since 1750, the number of scientific journals increased tenfold, reaching well over 100,000 by 1950, and straining the abilities of readers to stay abreast of the literature and retrieve earlier published information (1). Following World War II, science shifted away from strictly academically based, disciplinary research and toward mission-oriented, multidisciplinary approaches to solving social and technical problems. Thus, scientists had to draw needed information from a body of knowledge that was not only expanding at unprecedented rates, but compiled in increasingly diverse (and seemingly remote) sources.

This dilemma was quite pronounced in biomedicine: 19,000 biomedical journals were printed in 1950, well over 22,000 in 1970, and the boundaries that traditionally defined biomedicine as a field of practice and research were continually expanding to encompass new disciplines. Today, the field is so broad it defies attempts at delineation. In 1950, it was clear that traditional scientific publications were exhausting their effectiveness in communicating important advances and keeping readers abreast of the literature (1). Science turned to computers to accelerate the processing, storage, and retrieval of information, and thereby revolutionized all information-seeking activities.

Information Products and Services

The single most important source of biomedical information are primary publications—books, journals, technical reports, Government studies, patents, etc.—though the dramatic rise in their number often makes access to this literature difficult and confusing. Secondary publications—bibliographies, abstracts, catalogs, and indexes which facilitate access to the primary literature—are thus serving as an increasingly important element in the transfer of information.

Secondary services trace their historical roots to the learned and professional societies that flourished in the late 19th century, and which sought to fulfill the discipline-based information needs of their members in the basic sciences. In general, such services acquire and analyze primary publications, usually provide a bibliographic description of each document, indexing each according to an established system of content analyses, and, periodically, produce a product which provides subject access to the references for the documents indexed during a specific time period (112). Secondary services guide users to the primary source of needed data, rather than directly provide the information. By 1950, 3,000 abstract journals for the scientific literature were available in print form (1).

Until World War II, the need for interdisciplinary communication was met through a few broad-based professional organizations (e.g., the American Association for the Advancement of Science), the various academies (e.g., New York Academy of Sciences), and multidisciplinary publications (e.g., *Science* and *Nature).* When new scientific or technical areas arose, they were usually cross-disciplinary, like biochemistry. Information access needs could be met by establishing a new section of an abstracting and indexing service: witness the issuing of *Chemical Abstracts* by sub-area of chemistry, today including a section on biochemistry, and one on toxicology.

After World War II, the rise of Government-sponsored, mission-oriented research programs created a need for broader, multidisciplinary coverage of the scientific literature within a single secondary service. While professional (nonprofit) societies addressed this need through subspecialization, commercial organizations began offering selective, focused services relating to particular program areas that cut across disparate subject disciplines (e.g., the environment).

Technological Developments and Growth of On-Line Services

The secondary service adaptations after World War II were hard-pressed to meet the growing needs of the scientific community. In the late 1950's, producers of printed abstracts and bibliographic indexes turned to the new computer technologies to reduce costs by mechanizing the construction and publication of their products. The sophisticated application of computers to the processing of biomedical bibliographic information was pioneered by the National Library of Medicine (NLM), and by 1964 the first computerized biomedical bibliographic retrieval system, MEDLARS, was in use.

Most of the other machine-readable data bases in health-related fields, as well as in other subject matter fields, were originally built when the production of traditional indexing and abstracting journals switched to computer-driven photocomposition. It soon became clear that machine-readable data bases developed in the 1960's for the production of print copy could also be searched via computer: this development became the basis for on-line information retrieval services. These and subsequent developments were sponsored by both public and private enterprises. Government investment enhanced the potential of communication networks to attract other customers. The increased volume of business then lowered costs sufficiently to allow the industry to be cost effective and to grow.

All libraries have been affected by the electronic revolution, but none so much as those which serve the scientific, technological, and medical communities. Though these libraries have provided and will continue to provide traditional services and print products, the use of computerized services to support and enhance their services is rapidly growing. The growth of computerized services has depended on, and been driven by, the technical development and coordination of time-sharing computers, machine-readable data bases, fast-access disk storage devices, interactive retrieval programs, and low-cost terminals and telecommunications networks. Although this study focuses on data bases and their on-line services, other technological developments were necessary to the growth which we now see.

Machine-readable data bases, which began and continue as a more efficient method of preparing printed indexes, enhancing the value of primary publications, have grown in size and scope to compete with their printed counterparts, and form the basis of information transfer activities. A 1979 directory of data bases available on-line lists 528 separate entries, up from 301 cited 3 years earlier; 90 of these were relevant to biomedicine and health care (166). The number of requests for on-line searches grew from 700,000 in 1974 to over 4 million in 1979 and to 6 million in 1981. Over one-quarter of these were conducted on MEDLARS alone by users in 40 countries (1). Specific data on the services provided by commercial firms is considered proprietary and is thus unavailable for publication, though the industry as a whole is said to be enjoying a 20-percent annual growth rate (168).

As data bases have grown to be at least as, if not more, important than printed products as a source of bibliographic information, they have diverged from the publications they were developed to prepare. For example, a number of journals in the special interest areas of dentistry, nursing, and population science are indexed for NLM's major data base MEDLINE. However, these citations do not appear in *Index Medicus*. On-line products are beginning to contain more information than their printed counterparts, clearly the result of the economics of the process of information transfer to secondary sources. It is simply less expensive to add information to a data tape than to a printed product (112).

Appendix E.—MEDLINE: Technical Processes

Introduction

The processes involved in the creation and use of the data base MEDLINE were discussed in chapter 2. Four aspects of these processes—literature selection, the development of medical subject headings (MeSH), indexing, and searching—are described here in more detail. The same search process is employed for all MEDLARS data bases.

Literature Selection

The data tape used for the printing of *Index Medicus* is used as the source of data for MEDLINE. Thus, references to articles selected for *Index Medicus* are incorporated into MEDLINE. In selecting journals for *Index Medicus*/MEDLINE, the National Library of Medicine (NLM or the Library) receives the advice of seven to nine outside consultants who are acknowledged specialists in their fields. They are selected by the staff with the approval of the Director of NLM. Health educators, researchers, librarians and editors of medical and scientific journals have often been advisors. In 1981, in addition to representatives from these fields, the set of consultants included the vice-chairman of a health planning commission, who was formerly commissioner of a State health department. His appointment may indicate NLM's interest in health services research and delivery. Consultants are appointed on an ad hoc basis and serve no set term. They meet as a group three or four times a year, and the chief of the MeSH section/editor of *Index Medicus* acts as chairperson.

Journals considered for inclusion in *Index Medicus*/MEDLINE are suggested to the editor by publishers, the selection and acquisition staff of NLM, and, occasionally, by users. Several issues of each journal title are reviewed and discussed at length by the consultants.

In the selection process, the consultants consider the scientific merit of the publications, and their relevance to NLM's objectives. Other than these general standards, and a scope and coverage manual for the selection of materials for the Library as a whole, there are no formal selection criteria. NLM believes that the diversity of material included in *Index Medicus*/MEDLINE precludes the establishment of a set of criteria that would be relevant for all the categories included in the *Index*.

After each consultant rates each serial on a scale from 1 to 5, the chairperson averages the scores and ranks the serials according to their average ratings. With a score of 2.5, the serial is eligible for, but not assured, inclusion in the *Index*. In comparison to the number contained in NLM's collection, the number of serials indexed for *Index Medicus*/MEDLINE is very small. Only 2,664 of the more than 20,000 serials collected by the Library were indexed in 1980, an increase of slightly over 100 in the past 2 years. However, many of the serials are directories, annual reports, and newsletters. The number of articles indexed in 1980 was restricted to 273,750.

While management considerations (budget, number of positions, scope of the data base) have been important in delimiting the size of the universe of publications indexed for *Index Medicus*/MEDLINE, the question of quality control has dominated the restrictions on its size. The journal literature of the health sciences exhibits a high degree of repetition and a mixed degree of scientific excellence. The inclusion in or exclusion from *Index Medicus*/MEDLINE represents an effort to maintain a standard of scientific quality (3). With the purchase of the new IBM 370/168 computer, NLM expects to increase the number of articles indexed, and has requested that users, particularly foreign centers, update their equipment in the coming year to accommodate the change.

A similar mechanism is used to review journals that are being considered for deletion. In 1981, NLM added 145 new serials to *Index Medicus*/MEDLINE and deleted 142, 30 because they were no longer published. The consultants periodically review all the serials in an area, such as obstetrics-gynecology, and add or delete titles as appropriate.

As noted earlier, *Index Medicus*/MEDLINE has contained a small number of references to nonjournal literature such as monographs and proceedings of biomedical meetings. Because of the difficulty in selecting this type of material, the journal selection consultants recommended that this practice be discontinued; in 1981, it was. However, there is one type of document that is not covered in *Index Medicus*/MEDLINE and receives poor attention in NLM's entire collection, and that is the "fugitive" literature (see ch. 3).

Medical Subject Headings (MeSH)

The terms that are most commonly used by the authors of English-language literature are used in the MeSH vocabulary. New terms for the MeSH vocabulary are suggested by the literature itself, MEDLINE users, professional associations, staff indexers, biomedical scientists, and special, NLM-concerned committees. With few exceptions, MeSH tends to follow the literature rather than to lead it. One exception occurred when the American Psychiatric Association recently revised its classification of psychiatric disorders; MeSH was changed accordingly, even before the new classification had come into common usage. Similarly, NLM revised the MeSH terminology on the neurological aspects of speech disorders last year, with the assistance of experts in the field.

MeSH is arranged alphabetically and categorically. The 15 categories are further subdivided, and arranged in a hierarchical manner to show relationships between broader and narrower terms. The hierarchical structure and the indexing principles of specificity (i.e., indexing toward the most specific concept discussed) permits what is termed an "explode" capability during the search process. At the command of "explode," the system searches for all subcategories of a more general concept. For instance, if one were to search for the effect which a certain group of drugs, such as tranquilizing agents, had on animals, the "explode" feature would allow the searcher to specify only the phrase, "tranquilizing agents," and references indexed with the general term "tranquilizing agents," or specific agents would be retrieved.

Because the MeSH vocabulary is used for *Index Medicus*, for MEDLINE, and for cataloging, it has inherent problems; what is optimum for one may not be optimum for the other. For example, the *Index Medicus* user usually has a general interest in a subject and wants to search in fewer places than does the MEDLINE user who may have a specific topic of interest. Thus, the 14,000 MeSH descriptors are divided into 9,000 major and 5,000 minor descriptors. Whenever an indexer assigns a specific minor descriptor for the on-line searcher, the computer adds an appropriate predetermined, more general major descriptor under which the citation may appear in *Index Medicus*.

Indexing

There are prescribed qualifications for indexers who assign headings from MeSH to articles. In 1981, NLM had 21 full-time and one part-time indexers/revisers on its staff. Almost all of them had undergraduate training in the biological sciences; some held masters and doctoral degrees. One or two of the foreign language experts have had the science requirement waived. As noted in chapter 2, NLM and some foreign centers contract out some of the indexing to U.S. commercial firms. The commercial contractors are required to hire individuals with biomedical backgrounds acquired either through formal education or comparable experience.

All domestic indexers must take an NLM-operated 2-week formal training course. After completing the course, each new indexer is assigned to a reviser, who continues the training in an individual one-on-one setting. This one-to-one relationship lasts anywhere from 2 to 6 months depending on the reviser's estimate of the new indexer's ability. The training of foreign indexers has varied over time. In the late 1960's and early 1970's, many foreign centers sent their staff to NLM for training. Later, NLM sent experienced indexers abroad to some of the centers, to train the respective centers' staff. For the most part, new foreign indexers now receive their training from indexers previously trained at or by NLM.

NLM continually provides training to indexers. In the United States and abroad, indexers are in weekly or monthly contact by telephone or mail with their revisers. There is, as well, a day-long training session for indexers in the United States each fall, when the new MeSH is published. NLM sends orientation packages to indexers (usually from foreign centers) who do not attend the session. In addition, seminars with the revisers are held periodically in order to increase consistency of indexing. Technical notes and memoranda are published, generally monthly, to increase indexing consistency.

NLM staff indexers have specific performance standards to fulfill. They are expected to index four articles of medium difficulty per hour and five articles of lesser difficulty per hour. Revisers, who are highly qualified indexers, are responsible for reviewing and revising 15 articles per hour that are indexed by new indexers, and scanning 25 articles per hour that are indexed by more experienced indexers. Unlike NLM staff indexers, contract indexers are not held to index a set number of articles per hour, but are paid for each article indexed.

NLM exercises a high degree of control over the quality of the indexing by means of a sequence of computerized validation routines, by proofreading (which is performed at a number of stages in the flow of material through the indexing section), by controlling the qualifications and the training of the indexers, and by the use of revisers. In addition, NLM provides the commercial contractors with NLM indexing tools, such as MeSH, and requires the firms to supply the indexers with dictionaries, textbooks, and other aids.

Despite their qualifications and the training they receive, indexers make errors: a few find their way into *Index Medicus* and MEDLINE. Library personnel ascribe this to the complexity of the indexing processes and policies and to what they perceive as the inconsistencies of MeSH. There are, in particular, subtle differences between some of the subheadings, which may cause the indexers to be inconsistent in assigning them to MeSH. For example, the indexer may perceive the difference in the subheadings of metabolism or physiology as being so slight that the choice between them may be arbitrary in some instances. A computer program checks all indexing terms on the citation forms against MeSH, identifying spelling errors, ineligible subject headings/subheading combinations, and non-MeSH terms for correction.

In late 1981, an average of 69 days passed between the time a top priority journal was received at NLM and its entry into MEDLINE. Other journals took even longer. Since *Index Medicus* and MEDLINE are updated monthly, it would take the minimum of 30 days to process a journal. Concerned with the flow-through time, NLM is currently planning to partially computerize the indexing process.

Searching and Retrieval

Searchers formulate a search on the basis of knowledge of indexing principles, such as specificity, the use of subheadings, MeSH, and Boolean logic, and transmits the search statement to the computer via a keyboard terminal. (Searches can also be formulated on-line and then stored in the computer for reference and later use.) The computer searches the data bases and produces an individualized bibliography.

On-line search requests are generally broken down by trained information specialists into concepts that can be translated into MeSH terms and retrieved by querying the computer using one MeSH term or a combination of MeSH terms combined according to Boolean logic, a system that uses the connectives "and," "or," and "not" to express relationships between concepts. For example, to obtain every article discussing either potassium or cyanide, a searcher would ask for "potassium cyanide." To obtain only discussions including both poisons, a searcher would ask for "potassium *and* cyanide." To obtain articles dealing with potassium only when cyanide does not also appear, a searcher would ask for "potassium *and not* cyanide."

For new or abstract concepts that cannot be found in MeSH, the searcher uses text word searching. A text word search usually results in fewer, more specific citations than a search using MeSH. It also allows for retrieval of very new concepts that MeSH may not have incorporated into its vocabulary, and allows for the occasional user who may not know medical terminology to use the system. Unlike MeSH, text word searching does not have the "explode" feature and requires searching for a concept under all its possible expressions.

Appendix F.—Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science: Principles and Recommendations

The principles and recommendations of the Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science presented below were drawn from the 1981 task force report, *Public Sector/Private Sector Interaction in Providing Information Services* (97).

Governmental Leadership

Principle 1.—The Federal Government should take a leadership role in creating a framework that would facilitate the development and foster use of information products and services.

The recommendations related to this principle specify several areas in which Government can provide leadership: enhancing the competitive forces of the market-place; affirming the application of the first amendment; providing legislative consistency; using efficient technologies; supporting education, research, and data collection in this field.

Recommendation No. 1.—Provide an environment that will enhance the competitive forces of the private sector, so that the market mechanisms can be effective in allocating resources in the use of information and in directing innovation into market-determined areas.

Recommendation No. 2.—Affirm the applicability of the first amendment to information and services.

Recommendation No. 3.—Encourage Congress to be consistent in the language used and in the application of principles relating to information products and services, such as those identified in this report, when it formulates legislation and when it exercises its oversight role.

Recommendation No. 4.—Encourage Government agencies to utilize the most efficient (information) technologies.

Recommendation No. 5.—Encourage the setting and use of voluntary standards that will not inhibit the further development of innovative information products and services.

Recommendation No. 6.—Encourage and support educational programs that provide the professional skills needed to further the development and use of information as an economic and social resource.

Recommendation No. 7.—Encourage and support both basic and applied research in library and information science.

Recommendation No. 8.—Encourage and support statistical programs and related research to provide the data needed to deal with information policy issues.

Recommendation No. 9.—Conduct a periodic economic assessment of the impact of Federal Government information products and services.

Recommendation No. 10.—Encourage Federal agencies to regard the dissemination of information, especially through the mechanisms of the private sector (both for profit and not for profit), as a high-priority responsibility.

Recommendation No. 11.—Identify and evaluate alternatives to existing Federal information dissemination mechanisms.

Recommendation No. 12.—Develop and support the use of libraries as active means for access to governmental information by the public.

Encouraging Private Sector Investment

Principle 2.—The Federal Government should establish and enforce policies and procedures that encourage, and do not discourage, investment by the private sector in the development and use of information products and services.

Six recommendations are presented as means for implementing this principle. They relate to encouragement of new development, reducing uncertainties, and reducing risks.

Recommendation No. 13.—Identify and eliminate legal and regulatory barriers to the introduction of new information products and services.

Recommendation No. 14.—Encourage private enterprise to "add value" to Government information (i.e., to repackage it, provide further processing services, and otherwise enhance the information so that it can be sold at a profit).

Recommendation No. 15.—Provide incentives to existing organizations, such as libraries and bookstores, that will encourage them to expand their activities in dissemination of governmentally distributable information.

Recommendation No. 16.—Establish procedures that will create a realistic opportunity for private sector involvement in the planning process for Government information activities.

Recommendation No. 17.—Involve the private sector in the process of formulating standards relating to Federal information activities.

Recommendation No. 18.—Create or improve mechanisms for ensuring that the actions of Government agencies, in developing information resources, products, and services, are consistent with the policies, goals, and long-range plans that are announced.

Government in the Marketplace

Principle 3.—The Federal Government should not provide information products and services in commerce except when there are compelling reasons to do so, and then only when it protects the private sector's every opportunity to assume the function(s) commercially.

The related recommendations are to be considered as integral parts of this principle, since they embody the procedures for determining that there indeed are "compelling reasons" for the Government to provide services in commerce:

Recommendation No. 19.—Announce plans sufficiently ahead of time to provide an opportunity for private sector involvement when a Government agency, for reasons it regards as compelling, should plan to develop and/or to market an information product or service.

Recommendation No. 20.—Review and approve, before implementation, any plans for the Government to develop and/or market an information product or service, the review to be carried out by an agency appropriate to the branch of Government (such as the Office of Management and Budget, General Accounting Office, Congressional Budget Office).

Recommendation No. 21.—Include an "information impact and cost analysis" as part of the process of review, evaluation, and approval of any plans for the Government to develop and/or to market an information product or service, the analysis to cover economic and social effects on potential private sector products and services, and benefits to the public.

Recommendation No. 22.—Review periodically to evaluate the desirability of a continuation of any information product or service as a governmental activity.

Recommendation No. 23.—Do not arbitrarily restrict the Federal Government from enhancement of information products and services, even if solely to meet the needs of constituencies outside the Government itself.

Government Use of Private Sector Information

Principle 4.—The Federal Government, when it uses, reproduces, or distributes information available from the private sector as part of an information resource, product, or service, must assure that the property rights of the private sector sources are adequately protected.

Availability of Government Information

Principle 5.—The Federal Government should make governmentally distributable information openly available in readily reproducible form, without any constraints on subsequent use.

Recommendation No. 24.—Announce the availability of governmentally distributable information and maintain one or more registers to help the public determine what governmentally distributable information is available.

Recommendation No. 25.—Deposit governmentally distributable information, in whatever form it may be available, at national and regional centers, including regional depository libraries, where it may be examined at no charge.

Recommendation No. 26.—Do not assert any Federal Government copyrights on information the Federal Government makes domestically available.

Pricing of Government Information

Principle 6.—The Federal Government should set pricing policies for distributing information products and services that reflect the true cost of access and/or reproduction, any specific prices to be subject to review by an independent authority.

Use of Private Sector Dissemination Means

Principle 7.—The Federal Government should actively use existing mechanisms, such as the libraries of the country, as primary channels for making governmentally distributable information available to the public.

Recommendation No. 27.—Use the Nation's libraries and nongovernmental information centers as means for distribution of governmentally distributable information instead of creating new governmental units or expanding existing ones.

Appendix G.—AGRICOLA and ERIC*

Introduction

The Federal Government uses a variety of methods to produce and distribute data bases. The approach adopted by the National Library of Medicine (NLM or the Library)—creating and providing access, directly and indirectly, to its product—has been discussed in some detail. The efforts of the National Agricultural Library (NAL) and the National Institute of Education (NIE) described below further illustrate avenues available to the Government to ensure that needed information reaches researchers, practitioners, and in some cases, the public in a timely fashion.

The National Agricultural Library and AGRICOLA

The Agricultural On-Line Access (AGRICOLA) is a family of data bases containing citations to journals, monographs, and U.S. Government reports on agriculture and related subjects, including food and nutrition, economics, law, rural sociology, and many of the basic sciences. Currently, AGRICOLA contains 1.75 million citations to articles published in approximately 6,000 serial titles dating from January 1970. Unlike NLM, NAL does not provide on-line services, but rather makes its data base available to purchasers at a charge slightly above the costs of reproducing the file on magnetic tape, plus a minimal use fee. AGRICOLA is available on-line at $18 to $40/hour through the commercial services, DIALOG Information Services, Inc. (DIALOG), Bibliographic Retrieval Services (BRS), and System Development Corp. (SDC) (see table 8, ch. 4).

Legislative Mandate

AGRICOLA was developed and is maintained by NAL, the cornerstone of the U.S. Department of Agriculture's (USDA) Technical Information Systems. NAL began as the Department of Agriculture Library in 1862, but was not established as a national library until 1962. The congressional act establishing USDA states that "the general designs and duties [of USDA] shall be to acquire and to diffuse among the people of the United States useful information on subjects connected with agriculture in the most general and comprehensive sense of that word," and that "it shall be the duty of the Commissioner of Agriculture to acquire and preserve in his Department all information concerning agriculture which he can obtain by means of books and correspondence." NAL's current primary functions include coordinating a national agricultural science information network and serving as the prime information resource for land-grant and other academic libraries across the Nation. NAL is the second largest U.S. Government library in existence, and use of the largest agricultural libraries in the world. It has holdings of over 1.7 million volumes.

Indexing and Coverage

NAL is primarily responsible for generating the AGRICOLA data base. The original thrust was to automate the manual indexing procedures of the Bibliography of Agriculture, published since 1942. NAL receives approximately 25,000 journals per year and indexes over 6,000 of these journals and other serials for the AGRICOLA data base. It spends $710,000 per year on journal acquisitions. Titles are selected in accordance with the published selection policies of NAL and on the recommendations of NAL's Board and staff. In addition, agreements with 56 foreign countries establish a quid pro quo exchange of agricultural publications—approximately one-half of the citations in AGRICOLA are to international works, and 70 percent of the NAL collection is obtained through exchange arrangements.

Unlike NLM's MEDLARS, which uses free text word searching and medical subject headings (MeSH), a controlled vocabulary, AGRICOLA uses only free text word searching, though NAL is currently studying the United Nations' AGRIS (International System for the Agricultural Sciences and Technology) vocabulary for future use. AGRICOLA is updated on a monthly basis; 130,000 records are added each year.

AGRICOLA covers the broad field of agriculture and related subjects as listed below:

- agricultural economics
- agricultural engineering
- agricultural products
- animal husbandry
- aquaculture
- botany
- chemistry
- ecology and environmental science
- energy in agriculture
- entomology
- fertilizers
- foods
- soils
- forestry
- human nutrition
- hydroponics
- pesticides
- plant sciences
- rural sociology
- water management

In addition to AGRICOLA, USDA compiles CRIS (Current Research Information Service), a data base similar to NLM's CANCERPROJ. CRIS contains project summaries for 95 percent of ongoing U.S. agricultural and forestry research efforts, in all over 25,000 summaries. CALS (Current Awareness Literature Searching System) is similar to SDILINE, providing regular screenings of the current literature available through AGRICOLA and eight other data bases leased from commercial vendors.

---

*Information in this appendix was supplied by the U.S. Department of Agriculture and the National Institute of Education.

CAB (Commonwealth Agricultural Bureau), a data base produced in the United Kingdom, covers 60 percent of AGRICOLA citations. There is also some overlap with BioAbstracts and MEDLINE, especially for articles in biology and veterinary medicine.

AGRICOLA is also distributed by NAL through the National Technical Information Service (NTIS) to private producers of printed bibliographies (e.g., *The Bibliography of Agriculture)* and dictionary catalogs, and to several foreign Governments with large agricultural research centers. NTIS has full responsibility for the duplication and sale of the AGRICOLA tapes to 24 customers, though it relies on NAL in assuring user satisfaction.

Users

Like MEDLARS users, AGRICOLA users are difficult to enumerate and describe. NAL assumes that most users are researchers and educators who obtain AGRICOLA's services through the Nation's 69 land-grant colleges. Greatest usage probably is by employees of USDA though the system is valuable for researchers working in many of the general and applied sciences, agricultural engineers, food inspectors, and State extension agents, as well as faculty members, students, and private sector employees.

Document Delivery

NAL fills between 250,000 and 300,000 individual requests for books and journal articles per year. Document delivery is indeed one of major activities. Like NLM's Regional Medical Library Program, NAL has established 24 regional document delivery centers, based in the land-grant agriculture colleges, using its own resources only when documents are unavailable elsewhere. Documents are made available without charge to USDA employees. Other requests for photocopied articles are charged $3 per 10 pages. The regional centers are reimbursed for services provided to USDA researchers.

Training

Though AGRICOLA is available only through commercial vendors, NAL offers training courses in searching its data base. Workshops last 5 days, include at least 8 hours of on-line time, and are provided without charge. NAL's courses are less restricted by demand than those offered by NLM for MEDLARS, and are open to information specialists in both the private and public sectors.

Two levels of training are offered: one for those with no prior experience with AGRICOLA, and a second that focuses on subject searching for searchers with 6 months to a year's experience. The commercial vendors offer training programs on their systems, but they are not specific to AGRICOLA.

National Institute of Education and ERIC

ERIC (Educational Resources Information Center) is an information service funded by NIE (Department of Education) that maintains a bibliographic control system for educational journals and research reports.

ERIC's beginnings resulted from a boom in education research in the mid-1960's that overwhelmed existing systems' abilities to catalog and announce research findings to practitioners and other researchers. The agency's collection and indexing activities are decentralized around the country in 16 clearinghouses representing different levels and areas in education (see table G-1). Like medicine and health care, many electic subdisciplines contribute to the knowledge base of education, often without communicating with each other. The series of ERIC clearinghouses allows professionals tied to these communities, and with the expertise needed to identify important research from each discipline, to identify and select materials for the ERIC data base.

Publications

ERIC's principal activity is the publication of two secondary journals, *Research in Education (RIE)* and *Current Index to Journals in Education (CIJE).* Begun in 1966, *RIE* is a monthly bibliographic index to the "fugitive" educational research literature, e.g. reports of federally funded research that are neither copyrighted nor available in refereed publications. Reports that the clearinghouses index for *RIE* are available on microfiche from ERIC. *CIJE* was first available in 1969 and is similar to *Index Medicus.* The clearinghouses, with the assistance of advisory boards, select articles from 750 journals to be indexed and abstracted for *CIJE.* NIE has no requirements for what a clearinghouse must include.

For both publications, clearinghouses build bibliographic records of journals and reports based on a single thesaurus of indexing descriptors. Every month each clearinghouse sends its records to the ERIC central facility in optically readable, printed form, where the records are transferred to magnetic tapes. The tapes

Table G-1.—ERIC Clearinghouses and Operating Organizations, 1981

ERIC clearinghouse for/on:	*Operating organization:*
Junior colleges	University of California at Los Angeles
Science, mathematics, and environmental education	Ohio State University
Social studies/social science education	Social Science Education Consortium
Adult, career, and vocational education	Ohio State University
Counseling and personnel services	University of Michigan
Educational management	University of Oregon
Elementary and early childhood education	University of Illinois
Handicapped and gifted children	Council for Exceptional Children
Higher education	George Washington University
Information resources	Syracuse University
Languages and linguistics	Center for Applied Linguistics
Reading and communication skills	National Council of Teachers of English
Rural education and small schools	New Mexico State University
Teacher education	American Association of Colleges for Teacher Education
Tests, measurements, and evaluation	Educational Testing Service
Urban education	Columbia University

SOURCE: National Institute of Education.

are then given to the Government Printing Office (GPO) which publishes *RIE* and provides (without charge) tapes to the commerical publisher of *CIJE*, Oryx Press. GPO also furnishes the ERIC tapes to parties wanting on-line access to the data base.

Collections

ERIC is handled through GPO rather than NTIS because educational research information is outside the normal scope of NTIS interests. Additionally, it is important to note that while the ERIC data base is available on-line, the primary purpose of its construction is the publication of *printed* indexes, not the provision of on-line retrieval services. Thus, the on-line data base is a "spin-off" of the publications, and is sold to interested parties, including commerical vendors, for a few hundred dollars. Because the cost is so low, many universities purchase the tapes to mount on their own computers, giving them direct access to ERIC without using a commercial service.

In addition to the printed journals *RIE* and *CIJE*, Oryx Press publishes the *Thesaurus of ERIC Descriptors* from the ERIC tapes. The *Thesaurus* is the controlled vocabulary of educational terms used to enter and describe documents in the data base. Together, the three documents comprise a "minimal" ERIC collection, and are available at an annual subscription rate of $136 for domestic users and $153 for foreign users.

An "intermediate" ERIC collection includes the three publications and the ERIC microfiche collection, containing documents announced and indexed in *RIE*. Subscribers may order the entire microfiche collection from the ERIC Document Reproduction Service for $2,000/year, or may purchase subsets of the collection. Currently, these are over 700 subscribers to the microfiche collection, including over 50 subscriptions to foreign users.

Like NAL, NIE does not offer on-line services. ERIC tapes are provided at cost of reproduction, and are available at over 500 libraries and information centers. For the most part, those libraries are served by the three major commercial vendors at rates of $15 to $35/hour. As noted, about 50 organizations have their own computer facilities, and purchase and mount the tapes directly for their own use. In July 1979, the entire ERIC data base, including retrospective files, consisted of seven *RIE* tapes and four *CIJE* tapes and cost $875. Thus, to purchase a "complete" ERIC collection, including *RIE* and the microfiche collection retrospect to 1966 and *CIJE* to 1969, an organization would spend approximately $20,000, and contact ERIC again only for annual updates. A number of directories and indexes tools are available, often at no cost, to enhance any level of an ERIC collection.

Users

Providers of ERIC resources are called "access points" 52 percent of which were affiliated with institutes of higher education.[1] (Only discrete locations were counted; a single organization may have multiple access points.) There are an estimated 2.7 million "usage contacts" with the ERIC resources each year that prompt the identification of 32 million bibliographic records. Users spend 10.5 million hours annually with ERIC.

The number of on-line searches conducted on ERIC is growing. In 1980, 200,000 were performed, most (95 percent) by professional searchers and most often (65 percent) in academic settings. Of these, 50 to 60 percent were conducted for full-time students. Indeed, the bulk of ERIC "usage contacts" are from students and counselors, and for teaching and training educators. NIE sees the need to expose educational practitioners to the value of ERIC as its most pressing policy problem.

---

[1] National Institute of Education, personal communication, 1981.

NIE does not anticipate on-line searches of ERIC growing to replace the importance of its publications, *RIE* and *CIJE*. Yet ERIC is the most frequently searched data base offered by the three commerical vendors, BRS, SDC, and DIALOG. It is the major source of bibliographic information for the social sciences, as well as education. A layman's guide to using ERIC is under development to encourage utilization by practitioners, though only a modicum of training is funded through the clearinghouses, and that usually as part of students' coursework. No Federal funds are spent marketing ERIC or its services. The Government has taken a "low-profile approach" to developing a public awareness of the data base; users have discovered it on their own.

The Federal Government spends $5.5 million a year for the development of ERIC. NIE's budget has remained at the same level for several years—necessitating, due to inflation and higher production costs, cutbacks in the amount of relevant literature, especially journals, captured for the data base. The ERIC clearinghouses also receive support from their organizational affiliates.

Private Efforts

Wilson Publishing produces the *Education Index*, a monthly guide similar to *CIJE*. The two journals have some overlap; *Education Index* captures more of the topically relevant literature in publications outside traditional education disciplines, but is not so sophisticated a system for bibliographic control. It does not use keyword indexing, nor is it available on-line. Most educational libraries, however, have both ERIC and *Education Index*.

Appendix H.—Future Information Technologies: Implications for Biomedical Retrieval Systems*

Introduction

Electronic computation used to retrieve information from large data bases is an evolving technology, still only in its early stages. Its course can only be dimly perceived at this time. In the relatively near future, very little of today's techniques may have any more than historical interest.

Today's biomedical data base retrieval systems are primarily based on remote time-sharing computers, which means that the processing time of the computer is shared among several completely independent activities. Each user is unaware that there are other simultaneous users of the system, as the processor spends only fractions of seconds with each activity in turn before proceeding to the next. The computers compare indexed citations against a list of search terms. These terms are typed at remote terminals by researchers or trained computer search analysts. Terminals are connected to time-sharing computers via conventional telephone systems, specialized data networks, or both.

Projections for the next few decades indicate that changes are going to be even more radical than that of the past few decades. Machines are getting faster and more powerful. Telecommunications links are getting faster and capable of more sophisticated processing. Both are getting cheaper, but, relative to telecommunications, computation tends to become more cost effective at each technological jump.

Tomorrow's data base technologies may range over a wide variety of systems more powerful than the current methods. Full-text retrieval of articles, computer-aided searching of data, and other possibilities are being demonstrated today in experiments; some of these are expected to have an impact in the near-term future. The physical location where information is processed may alter within the next few years. High-speed data networks are expected to connect microcomputers sitting on researchers' desks; and it may prove more cost effective to move massive amounts of data processing locally than on distant time-sharing computers.

The present information policy issues are the result of usage patterns that have evolved over the past decade and may change greatly in the next few years as new technology based on distributed data processing, particularly the personal microcomputer, is applied to data base access. This observation is not meant to minimize the current difficulties. The present problems are important for the information industries today, and they are also important because what is done about them by policymakers now will affect the information systems of the future.

Past Application of Information Technology to Biomedical Bibliographic Retrieval Systems

Computers were first applied to information storage and retrieval systems in the late 1950's when a number of systems were developed in the United States by the Armed Services Technical Information Agency (now the Defense Technical Information Center), National Ordnance Laboratory, National Aeronautics and Space Administration, and the National Library of Medicine (NLM or the Library). The system developed by NLM, MEDLARS, was the largest of these both in terms of the size of its files and in terms of its user population. When MEDLARS was implemented, it was the first large computerized retrieval system to be made widely available without security and other restrictions.

The characteristics of MEDLARS are typical of the information systems that became available during the 1960's. The systems were operated in an off-line batch processing mode. This technique of batch processsing required data to be brought to a data center where they were punched onto cards, and then were subsequently read into the computer. The desired program, usually stored on magnetic tape, was also selected and entered from the tape into the computer memory. The program acted on the data and the results of the processing were output via a printer, or sometimes stored on a second deck of cards for further processing. The user of the data then came to the data center to collect the results. This form of processing worked quite well for applications that occurred at periodic intervals.

MEDLARS was originally designed not as a retrospective search system for bibliographic material, but as a publication system to produce the printed index, *Index Medicus*. The computer was used to manipulate bibliographic records in machine-readable form. It was used to check for errors, perform sorting and formatting, and interface directly with photocomposition equipment. For this, it was necessary to put the index records into machine-readable form and then update it, say monthly. Once this was done and the index produced, for example monthly, the machine-readable data base was available for further exploitation.

---

*This appendix is based on papers prepared for OTA by Richard Solomon, of the Research Program on Communications Policy of the Massachusetts Institute of Technology, Cambridge, Mass., and by Jose Marie Griffiths, of King Research Associates, Rockville, Md.

Such data bases could be used in a number of different ways. First, they could be used to conduct retrospective searches. These are searches through a body of recorded literature to find items on a specifically defined topic. Another application was for selective dissemination of information (SDI). In SDI, the current interests of users or groups of users are defined in interest profiles which are stored in machine-readable form. After the data base has been updated the additions are checked against the interest profiles. Matching records are printed out and sent to the appropriate users, thereby enabling users to keep up-to-date with the literature of their field of interest on a regular basis.

Computerized information systems developed in the 1960's offered significant advantages over their printed counterparts. These advantages are outlined by Lancaster as (79):

1. The possibility through batch processing of conducting many searches at the same time.
2. The ability to provide several access points to a document extremely economically.
3. The ability to handle complex searches involving large numbers of terms in complex relationships.
4. The ability to generate output in the form of a printed bibliography, and even to produce high-quality publications by interfacing the retrieval system with a photocomposition device. Output can also be made directly to microfilm.
5. The ability to collect, on a regular basis and essentially as a byproduct of normal systems operation, management data on how and how much the system is used.
6. The ability to produce many outputs and services from a single input operation. MEDLARS tapes, for example, although produced as a result of one indexing operation and one procedure for reducing the index records to machine-readable form, can be used to generate a general printed index, specialized bibliographies, retrospective search, and SDI searches.
7. The data base, once captured in machine-readable form, can be duplicated simply and cheaply; it is easily shipped around and thus can be used in the provision of information services by a number of different centers. This is perhaps the most important advantage of all. The growth of machine-readable data bases has had a dramatic impact on the provision of information services in the last 20 years.

Although the computer offered many advantages in information handling activities, the off-line batch processing systems have considerable disadvantages. Most significant is the delay in obtaining results, and if the search was unsuccessful on the first run, then it may have to be modified and reprocessed. In the same way, there is no facility for browsing through the literature. There was clearly room for further development of such systems.

In the early 1950's, it occurred to a number of people that the computer could be useful as a tool for researchers whose problems were too small to justify the initiation of a formal batch operation. Teletype terminals were sited in laboratories at the Massachussetts Institute of Technology, and these enabled researchers, each from his or her own teletype, to program the computer remotely. The computer would process each program in turn and return the results to the appropriate user's teletype as printout. This development marked the introduction of the on-line systems we are familiar with today.

The term "on-line" refers to the searcher's being in direct communication with the system he or she wishes to use. The user provides input to the system and it, in turn, reports back to the user. The user then makes decisions based on that report and provides further input. For this reason the system is also referred to as being "interactive" or "conversational." The development that made such systems feasible was time-sharing.

All these activities were carried out on mainframe computers and, to a large extent, still are. However, the late 1960's saw the introduction of minicomputers. Minicomputers were initially developed as low-cost computers with a minimum of processing power and memory. They were used primarily for control processing. It was soon recognized that minicomputers had a potential use in a number of unrelated areas and that they should be adaptable to the requirements of each. This resulted in the development of a number of general purpose minicomputers and the "unbundling" of software. Previously, machines were sold for specific applications with software provided as part of the overall system package—all "bundled in" together.

The advances in computer technology that have had a significant impact on information handling can be separated into two categories—processing and storage.

Processing

The first computer ENIAC (Electronic Number Integrator and Computer) was completed in 1946 at the University of Pennsylvania for the U.S. Army. It consisted of about 18,000 vacuum tubes and was very large, requiring a room 60 ft by 25 ft to contain it and weighing more than 30 tons. This computer (and others based on vacuum tube technology—later known as "first generation computers")—had high power consumption, had failures every few hours or so, and required cooling plants often as large and complex as the computer itself.

Two years later, the first transistor was developed at the Bell Telephone Laboratories. Transistors, while comparable to vacuum tubes in terms of the tasks they could perform, required much less electrical power, generated very little heat, and were more reliable. They became the basic component of "second generation computers."

Early efforts to miniaturize electronic components were not motivated by computer designers and engineers. Various satellite and missile projects, however, called for complex electronic systems to be installed in equipment in which size, weight, and power requirements were severely constrained. Thus, the effort to miniaturize was promoted by military and aerospace agencies.

The latest type of miniaturization is the semiconductor integrated circuit. Several researchers saw that the characteristics of semiconductors such as silicon or germanium that had been exploited to make transistors might be further exploited. The physical composition of semiconductors contained equivalents of individual electronic components (resistors, capacitors, etc.), and by combining them with transistors in the same material complex circuits could be created.

The integrated circuit was the basic component of the "third generation computers." In the early 1970's integrated circuits were being produced with about 1,000 components. The first microprocessor was developed by Intel Corp. in 1971, was about one-fourth inch square, and carried the equivalent of 2,250 transistors. By 1976, large-scale integration (LSI) produced chips carrying over 30,000 components, and by 1980 very large-scale integration (VLSI) saw chips with over 1 million components.

The individual elements on the silicon chips are defined by a photographic process. The smallest component dimensions on the chips are currently 3 to 4 micrometers (1 micrometer = 1 millionth of a meter). It is expected that this will be reduced in the near future to less than 1 micrometer and that by the 1980's it will lie between 0.05 and 0.005 micrometers.

The overall effect of these advances in processor technologies on information handling has been the reduction in size, cost, and operating requirements of digital computing equipment. Prior to the late 1960's, equipment had been far too expensive to justify use for what were then considered peripheral applications of information handling. However, by the mid-1970's many information-handling organizations were able to justify the cost of purchasing computing equipment, either alone or as part of a cooperative. Many large libraries and information services have their own small-scale equipment, but the continued reduction in size and cost of such equipment is beginning to influence the small and special libraries in particular.

Storage

The technology of digital storage is probably the most rapidly changing sector in all of microelectronics. Over the past decade, operating speed and reliability have been increased by at least an order of magnitude at the same time as physical size, power consumption, and cost per bit of storage have been reduced by factors of up to 1,000. Similar improvements are predicted for the next decade before physical limitations are encountered.

The newest electronic memory systems have been made possible by modern semiconductor technology. In the 1950's and 1960's, electronic memories were arrays of cores (rings) of ferrite material a millimeter or less in diameter. Ferrite core memories have now been succeeded, on the whole, by semiconductor memories which provide faster data access, smaller size, and lower power consumption at significantly lower cost. The overall effect of advancing digital memory technologies on information handling is an increase in the capacity and reliability of information systems, faster and more efficient retrieval of information (although because of increased capacity of storage, searches are made through larger volumes of information, and the faster access may not reduce the retrieval times as far as the user is concerned), together with an accompanying decrease in cost.

Current Status of Information Technology

Computer Processors

Although significant advances have been made in computer processor technology, very few of the commercially available retrieval systems have taken advantage of them. To a large extent, this is understandable. During the 1960's and 1970's, considerable investments were made in the acquisition of computer hardware. The effective lifetime of a single computer processor is 7 to 10 years, so organizations are unwilling (and often unable) to replace a processor within its lifetime. This is the so-called "technology trap." Those who hesitate to invest in a rapidly changing technology because new, improved products are imminent never reach the point of acquisition, and those who do make a decision to acquire a product acquire one which is usually no longer state of the art by the time it is in operation within an organization.

A second factor in the lack of change in commercial bibliographic retrieval services is that the functional capabilities of minicomputers and microcomputers are not as great as with the larger mainframe machines. Over the next 5 years or so, this situation is likely to change. The larger end of the minicomputer range (the 32-bit machines)—the superminis or megaminis as they are sometimes called—will be able to compete with the large IBM, Univac, Burroughs, and similar machines, both in terms of system software (operating systems and system utilities) and number of users that can be simultaneously serviced.

Areas where smaller machines have been applied are in individual organizations and libraries. Minicomputers have been used successfully in libraries for many years now, primarily for automation of many of the day-to-day library operations—acquisitions, cataloging, and circulation control. As the costs of computer processors fell dramatically, a number of libraries began to acquire these low-cost machines for maintaining bibliographic files for relatively small collections. The retrieval programs available on such small machines are not as sophisticated as those for larger systems, nor is retrieval as fast, but they are quite adequate for the environment within which they are applied.

Storage Devices

Digital storage devices are probably the fastest moving area of microelectronics. In particular, two new types of digital storage devices are likely to have a significant impact on retrieval systems in the future: 1) video disks, and 2) charge-coupled devices (CCDs) and bubble memories.

There are two approaches to recording information on video disks (or optical disks). One is an analog method which yields a color television (TV) picture, and the other is a digital method which stores data in digital form. It is the latter type, in particular, that could revolutionize information handling in the not too distant future. The current capacity of such disks is of the order of 1011 bits. This means that the equivalent of the Library of Congress card catalog could potentially be stored on one side of a digitally encoded optical disk. However, this is unlikely to happen for several years yet.

The disks can be written once only. A laser is used to burn away a thin metallic film from the disk surface to create a hole about 1 micrometer (one ten-thousandth of a centimeter) in size. The presence of a hole signifies a 1, and the absence of a hole signifies a zero. The projected cost for the disk is $10, and copies could be made for almost the same amount. A disk can store about 10,000 books (the full text), making the projected cost per book 0.1 cents. The bulk of the cost (over the $10 disk cost) will be the cost of digitizing the textual material and the disk reader.

The other type of storage devices under development are electronic serial access memories, in which the stored bits of information circulate as if they were in an enclosed pipeline. Each bit that is stored is transferred sequentially through 64 or more storage locations. These memories are smaller and cheaper to produce than other forms of electronic memory, mainly because the circuitry can be simpler. They cannot, however, compete in terms of speed with other electronic memories. Their most attractive potential application is the replacement of tape and disk memories with a capacity of on-board storage of between 1 million and 10 million bits.

COMMUNICATIONS

Most data communications systems currently under development are based on the concept of packet switching. The rationale for packet switching is that the length of a message is very short in relation to the time it takes to establish a conventional communications connection, and that it is more economical to transmit that message stage by stage through the network, storing the message segments at each node in a network until a link is available for transmission to the next node. It is now possible to purchase services from a commercial packet switched communications network.

To some extent, the microprocessor counteracts the rapid growth of networks. The central idea of networking is to share both functions and data. As small computer systems have become more powerful and compact, they have also become more independent in function. People are now able to buy low-cost computers to solve most of their routine problems. The general trend is away from network designs based on functions shared over a network of computers towards data sharing across networks.

The major trend in telecommunications at present is the fundamental shift from analog to digital modes of transmission. The new transmission channels (based on optical fiber technology) will have enormous capacities. The shift involves all types of communications—voice, facsimile, computer transmissions, TV communications, microwave and satellite communications, and radio links. Digital circuits are less prone to interference and noise than analog ones, and their cost-performance ratio is constantly improving.

Communications may be expected to play an increasingly central role in information handling. Satellite-based communications systems offer a different approach to the present ground-based systems in terms of technology, but the services offered by the two systems should be similar. Satellite communications will probably develop as an extension to the ground-based systems, and will provide a more cost-effective solution in sparsely populated areas, for example.

INPUT/OUTPUT DEVICES

New and improved means of communicating with computers are pouring onto the marketplace at an ever-increasing rate. The standard teletype terminal and cathode ray tube terminals for interacting with computers are being replaced with sophisticated graphics devices, color screens, touch-sensitive screens, letter quality printers, laser printers, and so on. Many of these input/output devices themselves contain microcircuits, thereby providing the user with localized intelligence.

The key change in the mode of interaction with biomedical bibliographic retrieval systems has been in the use of graphic terminals (especially for displaying organic compounds in schematic form) and/or intelligent terminals for search development and storage on input, and bibliography editing, sorting, and merging on output. However, until the retrieval systems themselves change, the mode of user-system interaction is unlikely to change much at all.

MASS MEDIA COMMUNICATIONS

One of the growing applications of communications systems to information handling is the provision of electronic mail services. It is possible to convert any type of message into digitized for transmission. Naturally, different messages require varying numbers of bits for their representation, for example:

• high-quality color photograph	2 million bits
• newspaper-quality photograph	100,000 bits
• color television frame	1 million bits
• brief telephone voice message	1 million bits
• document page in facsimile form	200,000 bits
• document page in computer code	10,000 bits
• coded request for library document	200 bits

Recent additions to information services are those based on the broadcast systems. Two types of systems exist: Videotex (sometimes called Viewdata) and teletext. Although their services and technology differ, both are accessed through the adapted domestic TV set, distinguishing them from other home information utilities that require computer terminals.

Videotex is an interactive system linking computer data bases to the adapted television set through the switched telephone network. Canada, France, and the United Kingdom have developed systems of their own and are selling the technology to other nations. A variety of Videotex systems exist, but they all operate in basically the same way. To access the signals transmitted via the telephone network, users must have special decoders and a modem built-in or attached to their TV sets. To connect to the central data base the user must first dial the appropriate telephone number and place the receiver in the modem. When the connection has been accomplished successfully, an index page appears on the TV screen and users begin to search for the information they require by pressing numbered keys on their hand-held control panel (keypad). Instructions appear on the screen telling the user which keys to depress for particular types of data.

The central Videotex data base may contain an almost unlimited amount of information provided by sources ranging from local newspapers to travel agents, department stores, and libraries. The data are stored in "frames" or screenfuls and can be updated instantly. Several frames of information on the same topic comprise a "page" and may be accessed sequentially. To retrieve information from the data base, users employ a "tree structure" search method, starting with broad subject headings and narrowing down their choices until they arrive at the frame of information they require. Such a system is fairly limited in capability and has no cross-referencing.

Unlike Videotex, teletext is a noninteractive system linking the information provider to the home via regular or cable TV broadcast signals. Once again, only TV sets with special decoders are able to pick up teletext. Pages of information are broadcast one at a time in recurring cycles. To access them, users consult a contents page, then use the keypad to key in the numbers of the pages to be retrieved. The decoder then selects the appropriate pages when they cycle by, and the information is displayed on the TV screen.

Teletext's chief virtue is its ability to be updated continuously for a large viewing audience. It can provide users with the most current information on a range of subjects and is easily accessible. Because it is broadcast rather than telephone based, teletext is also less expensive than Videotex, which requires users to pay for the telephone service and for each individual frame accessed.

Teletext has disadvantages, the greatest being its limited data base size. To access information, users have to wait until the specific page they are seeking cycles by, giving the decoder time to read, decode, and display the data. The wait time becomes excessive when the data base exceeds about 100 frames. Thus, teletext is severely limited in the amount of information it can carry efficiently.

Developments in the areas of processing, memory and storage, and communications interact to produce some very sophisticated information systems. On the whole, these systems are more compact, less expensive, considerably more reliable, store larger quantities of information, and offer newer types of service than the traditional type of information service. Moreover, access to the systems from remote locations is simpler and more efficient than ever before. Can we cope with the volumes of stored information that are available to us?

It is interesting to note that with the emergence of new technologies there has been a fundamental shift in the concerns of the information profession. This shift can be expressed, in simple terms, as a change from the attitude—"How can we give the user more information to solve his problem?" to "How can we make sure that the information we give the user is accurate, reliable, and up-to-date?" The basic questions concerning the organizations, classification, indexing, and retrieval of information are regaining importance, having been overshadowed over the last few years by the concerns of automation of systems and services.

With increasing volumes of information being stored and transmitted in digital form, and access to the stores being effected increasingly from the home environment, attention must return to the traditional areas of information handling mentioned above. It is envisaged, therefore, that a number of investigations will be initiated in the near future to consider potential solution to the problems of selection and quality assurance.

Future Trends

There are several trends under way today that will change the way researchers access data bases within the next few years. They cannot be ignored, for unlike the pioneers in the first generation of information retrieval, today's pioneers are outside of the biomedical field—and the evolution of the technology will go on no matter what resolution is made of current problems with access to MEDLARS. These trends are summarized here, since detailed discussion is beyond the scope of this paper. Many of the problems that private vendors have with NLM's policies reflect tensions being created by these changes.

1) *The widespread use of microcomputers is expected in the future.* We are now in the age of desktop microcomputers that have the power of the large computers of about a decade ago. Within the next 2 years, desktop machines with the ability to make sophisticated searches of the entire MEDLARS data base will be for sale in computer stores. Many, if not most of MEDLARS users in the developed nations might have such machines as general-purpose office research tools just as they now have typewriters, calculators, and lab equipment.

The key to matching indexes to a citation or abstract data base lies more in the size of a computer's core memory and the amount of external data it can access rapidly than in its speed of operation. About ½ million to 1 million characters (1 megabyte) or more would be adequate for a stand-alone desktop machine for data base retrieval. Such machines have been on the market for the past 2 years, costing about $30,000 to $40,000. Their price should drop by half or more in the next few years. These are also precisely the type of machines that many nonscientists would use for general number crunching, and they may make central processing on time-sharing systems obsolescent.

2) *Timesharing, as a model for data base access is expected to be radically altered as distributed computing becomes more widespread.* In the 1970's, data bases were placed on large central computers to which many terminals had simultaneous access. That was called "time-sharing." MEDLARS data bases and most of today's electronic data bases reflect that technology, which is undergoing transformation. The coming crop of 16-bit and 32-bit microcomputers attached to software represents a several orders-of-magnitude advance in computing, while commensurate improvements in long-distance data communications, and thus in timesharing, during the next few years may be only one or two orders of magnitude beyond that of today.

With a microcomputer, a user would only have to slide in the data base mass storage device—perhaps a video disk of the data base without waiting for a time-sharing connection or paying time charges. The latest information, always a tiny percentage of the total data base, could still come via telecommunications lines, but not necessarily at that instant. Each evening, for instance, a microcomputer can be programed to call in for the latest citations, articles, or specialized data. The amount should be able to easily fit on the disk drives, which hold some 5 million to 20 million characters of data. There are now available for desktop machines at costs ranging from $2,500 to $6,000.

Any user will tell how constrained one can feel when trying to browse through an on-line data base and the meter is ticking; in fact, browsing and experimenting with searches is the antithesis of current data base networks, but would be encouraged by distributed procession on a user's own machine. Many other features would be possible—e.g., generating a private profile of interests, reuse of material often needed, retransmission of important information to colleagues, reformatting information to one's own needs or tastes.

3) *Computing costs are expected to drop faster than communicating costs.* This relationship has been true for the past 30 years and all indicators show that the trend will continue for the near future. While these technologies are linked—the cost of telecommunications drops because of computer-related and electronics advances—the decline of both do not follow uniform parallel lines. While the cost of computing has been dropping regularly and in half every 2 to 3 years, universal telecommunications advances tend to come in somewhat unpredictable spurts (for reasons that are beyond the scope of this study) and on the average at a slower rate.

As a result, the "time windows" are different for the introduction of computing and communications products. It appears that computing devices for local data base processing will enter the market about 2 to 3 years before the advanced telecommunications systems. These data base products are now available: the microcomputers already mentioned, and laser video disks capable of storing all of MEDLARS on one or two disks.

Laser disks, which can store digital data, are being manufactured commercially today for about $5 in small lot quantities. The master disk costs about $2,000, not including the cost of the original videotape. Converting computer data to a video pattern is no major obstacle. But problems of data integrity and software for search and access on such disks are still to be resolved.

If MEDLARS had 5,000 customers worldwide, each paying $500 per year for a complete updated data base to date (assuming two disks), then the cost of producing 10,000 disks would come to $50,000 + $2,000 + distribution. The revenue of $2,450,000 annually surely would cover distribution and a great deal of the cost of production of the data base. Another $500 per year could be charged for updating via floppy diskettes or some other system, and the user would still be saving a great deal of money, not to mention the added revenue to NLM or a private vendor providing the service. These figures are only meant as an illustration of the economics of the new medium of video disk plus microcomputers for data base access; the normal scientific marketing algorithms for pricing such a service might come out with more appropriate fees.

4) *The future telecommunications networks are expected to mix data, voice, and other traffic making it simple to connect distributed machines but also confusing national boundaries, property rights based on print, and what is information and what is communications.* Even Euronet is expected to disappear as new technology is introduced that will merge telephone, Telex, and data systems into "integrated services digital networks" (ISDN). With ISDN, virtually all local telephone connections will be able to handle voice as well as high-speed data; and the data links will connect directly to a packet-like network permitting instantaneous connections, the ability to transfer hundreds of pages per minute between computers or desktop workstations, and most likely distance-insensitive pricing for data. Most major cities, however, will offer some sort of ISDN service by the end of the decade both in Europe and the United States, if replacement projections for older telephone equipment continues at the current rate. In the United States, such networks should begin to emerge within the next 3 to 4 years, at first in industrial suburban and some rural areas.

5) *Copyright questions in the computerized information age will continue to be extremely complex.* International property rights are particularly complicated, since property rights are defined differently in different jurisdictions. Though U.S. copyright law forbids NLM from claiming copyright protection domestically, NLM places a copyright notice in its printed material claiming protection internationally, and claims contractural rights both domestically and internationally preventing computer centers from replicating its tapes.

Furthermore, applying property concepts derived for the printed press to computer-based scientific information necessary for worldwide health and biomedical research creates a need to reconcile some mutually exclusive goals: Do we want to minimize U.S. Government expenditures in regard to medical information at the risk of encouraging the biomedical data bases which may create a negative balance of trade in information? Do we want to restrict access by foreigners to Government originated publicly available information? Is it strategically effective to try to use this resource to trade for similar rights to foreign data bases? Do we want the originators of the literature to share in the proceeds from abstracting or retrieval or do we want to encourage the widest dissemination of information at the risk of discouraging future information providers? (Abstractors or bibliographers in the print media have never had to pay royalties to the authors of the material described.) How do we fix property rights if a computer does the abstracting? What rights do the authors of the abstracting program have? Countries are likely to answer these questions differently—once more making costs in different countries very different.

6) *Machine retrieval of biomedical information has only scratched the surface with on-line literature citations and abstracts; the technology is evolving to permit retrieval of full text of articles, raw data, and eventually automated search.* Precedents established for

the rather primitive technologies of today can hamper further developments in the future; the converse could also be true: imaginative solutions to the domestic public/private mix of MEDLARS data base vending may encourage new techniques and industries to develop in the United States, rather than elsewhere. Several complex issues come to bear here: 1) the ability to enforce a form of international copyright on U.S. Government-generated data; 2) the property rights of the originator and "publisher" of the cited material; and 3) the comparative economics of data bases development and input outside and inside of the United States.

Solutions to these problems are not going to be simple. Transnational as well as cross-industry agreements will be necessary to protect information property rights. In some cases, there will be conflict between societal goals, particularly where information important to policy and technological decisionmaking is involved. Designing laws that will withstand decades of change is going to take a great deal of skill, both in anticipating technology and in anticipating new applications. The effects of new interventions are often unpredictable.

Appendix I.—Issues in International Access to MEDLARS*

Introduction

Though the congressional request for this study did not specifically ask OTA to address issues in international access to MEDLARS and international health information policy issues, understanding these issues is essential to the development of sound public policy toward health information resources. Those issues are so complex that a comprehensive assessment of them is beyond the scope of this report. This appendix identifies and discusses selected relevant issues, to both illustrate this complexity and demonstrate the necessity of addressing the international aspects of U.S. information policy. The discussion is divided between developed and developing countries, though in theory and in the future issues relevant to developed countries may become applicable to the Third World.

Quid Pro Quo Arrangements

The National Library of Medicine's (NLM or the Library) activities have extended into the international arena since its earliest years. Its responsibilities to serve as a repository of biomedical literature necessitate that it actively acquire and preserve information from around the world. Indeed, fully two-thirds of the journals included in *Index Medicus* are published abroad.

The international nature of NLM's collection is reflected in its MEDLARS data bases, so that many countries were naturally interested in obtaining access when the system became operational in 1964. The Library established quid pro quo arrangements, first with Sweden and the United Kingdom, to provide foreign countries with the MEDLARS tapes or on-line access to the NLM computer. In exchange, participating countries provide indexing and other services and/or pay commercial firms in the United States to perform such services for NLM. NLM currently has quid pro quo agreements with 13 countries and one international organization, the Pan American Health Organization (PAHO).

NLM enters these bilateral agreements with a single MEDLARS center selected by each country as its national biomedical information resource. The MEDLARS centers must meet NLM-established technical requirements for personnel, equipment, and fiscal resources, and serve a user community sufficiently large for the support of extensive computerized services. Indexing services are either provided in-house by the foreign centers, or contracted out to American firms. Though no moneys are transferred between NLM and the foreign centers, these indexing services are valued by the Library at $500,000. The agreements allow foreign centers access by mounting the MEDLARS tapes on their own computers with or without NLM software, or directly accessing NLM's computer (see table I-1).

In the late 1960's, NLM and the Organization for Economic Cooperation and Development (OECD) discussed the possibility of establishing a single consortium to serve as a European MEDLARS center, but no agreement was reached. Had events in the late 1960's moved differently at OECD—and at the European telephone and telegraph administrations—we might have seen totally different patterns for MEDLARS access, by 1970 it was clear that transatlantic satellites and specialized terrestrial data networks were going to be an important force in data communications. With only one center in Europe, and one in the United States, national centers would not have developed and the issues of restrictions might have taken a different course today, but the concept was technically and politically premature. Communication linkages were poorly developed, and the European nations lacked a common will for cooperation (3). The incipient changes in technology which tend to track anything related to computers might have been much easier to introduce in 1981 if OECD had been foresighted in 1966. This has obvious implications for the decisions we may be making in the next few years about the next generation of U.S.-sponsored public data base system.

Table I-1.—Foreign Centers' Access to MEDLARS, March 1982

Tapes	Tapes/software	On-line NLM
Germany	Sweden	France
Japan	United Kingdom	South Africa
	Australia	Canada
	Pan American Health Organization	Mexico
		Colombia
		Kuwait
		Italy
		Switzerland

SOURCE: National Library of Medicine.

---

*This appendix is based on papers prepared for OTA by Richard Solomon of the Research Program in Communications Policy at the Massachusetts Institute of Technology and Tefko Saracevic of the School of Library Science and Information at Case Western Reserve University.

International Issues: Developed Nations

Restraint of Trade

Until recently American commercial data base vendors have felt that they were in competition with their own Government in trying to sell biomedical information overseas. This has arisen because of the complex relationship between health services and governments in general, not the least domestically. As it has become profitable to sell health information via remote computers, the question of the proper role, if any, for a governmental agency in the United States has been introduced. But other governments have different views on this subject. In general, medical care is a quasi-governmental concern in most developed countries, historically linked to concerns for public health responsibilities.

Quid pro quo arrangements have complicated the problem: their general pattern is that there be " . . . no transfer of monies between the participating country and NLM." NLM makes MEDLARS available, either through tapes or (when this became possible) on-line access to the NLM computer, and provides technical documentation and training. The participating country must meet technical criteria involving personnel, equipment, and fiscal resources and have a user community large enough to justify an extensive computerized service activity. The participating country then provides and/or funds the indexing of journals for input to the MEDLARS data base in return for access to the system. This concept is consistent with a policy adopted in 1966 by the Committee on Scientific and Technical Information (COSATI) of the Office of Science and Technology in the Executive Office of the President, which was that Federal information systems would be made available in return for some form of contribution (31).

NLM at first had only one licensing agreement with a U.S. commercial information service, Bibliographic Retrieval Services (BRS), to lease the MEDLARS data base tapes. This agreement limited the firm to vend access to the data bases to the United States and the U.S. territories, unless specifically authorized by the Library in writing. Upon the firm's request, NLM permitted BRS to provide information services to countries outside the United States for which the Library did not have bilateral agreements in December 1978. An amendment to the licensing agreement was executed in March 1979, which, in effect, made it possible for foreign users of MEDLARS to connect to other sources if they were willing to take the risks, and make the effort, of establishing such connections since transnational telecommunications links were readily available, though at some cost.

In spring of 1981, the policy was modified and the new contracts that NLM has with commercial information services do not contain a national exclusivity clause. However, new contradictions and unenforceable requirements have now been introduced in regard to Communist countries by the Department of Commerce. The standard contract has recently been extended to prevent resale of data bases (from either U.S. vendors or foreign centers) to the Soviet Union or the People's Republic of China without NLM's permission, but not to other Communist nations.

This approach is ineffective, since NLM has no mechanism to investigate all their potential users that are given authorization user codes. Merely restriction by location is ineffective for electronic information transfer. All a potential user in those nations would have to do is obtain an authorized use code in an authorized nation and access the MEDLARS data bases through any number of public networks. In fact, this is a common pattern for information transfer to Soviet-bloc nations. Such rules are almost as unenforceable as would be one that told book buyers that they may not resell a book at second hand to a Communist buyer.

However, the major issue of restraint of trade due to geographic restrictions appears to be resolved for the most part. The private information services still are concerned that all licensees of MEDLARS data base tapes pay exactly the same amount. The current price structure and terms are the same for these with whom NLM has international bilateral agreements as for the U.S. domestic licensees who are commercial organizations. Nonetheless, some private sector firms state that quid pro quo services are not equivalent to the transfer of money. NLM notes that "the fees are identical for a bilateral center or a license" (110). In addition, the latter saves U.S. tax dollars and brings dollars to U.S. businesses, who frequently provide the actual indexing under contract with the foreign country (110).

Transborder Data Flow—National Restrictions

The ability to access MEDLARS in Europe via means other than national data base centers is possible because the growth of specialized data telecommunications networks. During the past few years, the European telephone administrations have instituted several different types of networks optimized for the connection of terminals to time-sharing computers. In addition most of the European Post, Telegraph, and Telephone Administrations (PTTs) have joined in a continentwide consortium to provide data services, called Euronet. Most important to the users of data base computers, such as those which offer MEDLINE, the cost of service for access is not based on distance.

Instead, charges are normally related to the connection time and the amount (measured in bits) of data transmitted, plus a connection charge. As a result, a use often can access a distant time-sharing system as easily, and at the same telecommunications cost, as a nearby machine.

These European systems are similar to the "packet nets" in the United States, but the state-monopoly PTTs have chosen not to pass all the savings onto the end user. On packet systems, data is only transmitted between terminals and computers when a key is pressed or a letter is received, instead of maintaining a continuous connection between two points, as on an ordinary telephone or Telex circuit. This data traffic travels in relatively small packets and is switched or routed via a complex network of computers that may process the data as well as transmit it. Though this technique shows distinct economies in line utilization, since most traffic on time-sharing systems tends to be in small "bursts," traffic shifts from traditional Telex systems represents a threat to the PTT's revenue base.

Technically, a worldwide system of packet data networks could equalize charges to virtually any point on the globe, since packet networks have the additional advantage of satellite technology with computer switching. But in practice other factors have intervened, so that artificial pricing barriers to uniform-cost data transmission exist between Europe and North America.

Restrictions on access are set by each nation, usually through telecommunications tariffs, or by general agreement among the user community. However, due to open access to telecommunications links across European borders, backed by agreement among the European nations themselves, enforcement of such tariff restrictions is purely fiction. Enforcement by NLM or its vendors has only been through password controls, which are extremely difficult to monitor.

In many countries, inducements are offered to process data locally. It should not be surprising that a nationally subsidized medical community could be encouraged by another government agency to do its bibliographic searching on domestic machines. In some cases, MEDLARS services are offered at rates lower than by NLM itself. Despite subsidies, in general these institutional goals to localize data processing are often in conflict with the goals of the data base user. The user prefers convenience in accessing systems, the opportunity to access competing services to maximize retrieval possibilities, and to minimize total costs. The telephone carriers, some of whom also dominate time-sharing or data processing in their countries, want to maximize revenue and are not interested in competitive services.

Other subtle restrictions imposed by the foreign nations on the use of the NLM data base overseas are designed to prevent data from being processed outside a nation's borders. These rules are somewhat indirect, and rarely overtly stated in terms of computer processing. One rule concerns transmission lines used for data which are tariffed at a higher price than if used for voice, so a user may not send data over an ordinary telephone line. The other rule is that the device that attaches a digital computer terminal or computer to the analog telephone line (called a "modem" for modulator-demodulator) must be obtained from the telephone administration. Where the telephone administration is a powerful government agency, such as in West Germany, enforcement of such rules severely restricts attaching terminals and computers to the telephone system.

To overcome some of the problems of access to U.S. data bases, some American firms offer the full MEDLARS data base via time-sharing systems on Euronet. Due to PTT tariff barriers, there is a tradeoff between transatlantic data links and European data processing in favor of the latter for certain services. BRS has licensed Datastar in Geneva to offer MEDLARS, using the proprietary BRS search system, and Lockheed similarly will mount MEDLARS tapes on a computer in London connected to Euronet. Separate copies of NLM's tapes are purchased for these European machines, since the NLM contracts prevent copying of the tapes (except for normal time-sharing computer operations at a single site). (See section below on copyright.)

Copyright

The use of copyright to protect various forms of property or artistic rights is not an abstract application of property law, but has been traditionally linked to the technologies of creation, reproduction, and distribution of literary, artistic, or commercial works. The radical changes in information processing technology which will affect international data base access become even more confusing in the sphere of copyright law, especially when applied to U.S. Government-owned materials such as MEDLARS.

Technological changes in the retrieval of information from computerized data bases have an important impact on international copyright issues. The computer is not like the printing press; it does not even work like an office copier. To be sure, there are some superficial resemblances to "publishing" or "copying" using computers, especially since some computer systems use such printing devices for paper output. But this imitation of publishing is merely a transitional phase. The computer manipulates information in ways that were never dreamt of by the framers of existing laws on copyright and the property rights of authors.

Because computers perform the functions of storage and dissemination of information in a radically different manner, discussion of issues of international copyright protection must recognize that existing technological concepts cannot serve as a framework for the future. Even current concepts of data processing are rapidly becoming obsolete; these are not dreams, there are working models of future systems now in place in many institutions and research organizations.

There is a radical difference in concepts of copyright based on the printing press, or on similar devices that slowly impress exact copies of an artistic or literary creation, and property rights based on the use of computation-based techniques to store and replicate information. For the printing technologies, copying is the penultimate step before distribution or storage; while the use of computers to perform parallel functions requires "copying" at every step in the process, from creation of the work to its final application. Since, in the coming decades, most useful information will be machine-readable, automation of reproduction has altered the essence of "copying," as well as "authorship" and the meaning of original work. Even today, magnetic and optical methods, and computerized techniques of data manipulation, make it possible to automatically reproduce and save virtually anything humans can see.

Attempts to enforce rules that were designed for a different age may make the existing copyright conventions and laws ineffective, and perhaps counterproductive. For example, with modern electronic technologies, pirate editions of the MEDLARS data base could be produced by Third World or Soviet-bloc nations for distribution on some mass storage device, perhaps by scanning the printed data base with a laser character reader and placing the resultant data on a video disk; justification for such activity could be made on the basis that public health-related materials should have as low a price as possible, and copyright enforcement is unfair to poorer nations.

Of course, despite high motives, distribution of the pirated data base would be open to all comers. More important, quality control by the original producer would be lacking. Essentially, what the pirate would have accomplished is a shift of income from the MEDLARS producer to the pirate. Since he would not have to produce the original data, the pirate would have lower overhead. But this form of cream-skimming simply increases the costs to the legitimate data base subscriber. Electronic publishing contains many variations of this theme, with other less invidious schemes for capturing and replicating information.

Furthermore, the NLM contracts with licensees contain provisions regarding replication of the data base that make analogies to printing in the mounting of computer tapes. The vendors will deny that they copy the tapes for any purpose when they mount them on their machines, but this is a contradiction in terms as any programer knows. There is no way to read a tape without copying it in some fashion. Yet, despite this fact, data base vendors feel they would have to purchase a second set of tapes, if they maintained an overseas data center. This would be unnecessary if the vendors operated a distributed system with a link between the United States and the foreign center, but their interpretation of the NLM contract assumes that such an operation would be prohibited, and is not even considered despite potential customer savings and other efficiencies.

We are witnessing the reactions to early phases of this new technology in the rules and laws promulgated which restrict certain information flows. This came about as telecommunications networks were established that made large-scale data transfers much easier between computers. This has made property rights in information more difficult to protect. It is in the interest of U.S. data base vendors, the U.S. Government included, to get involved in the protocol discussions for these data communications networks so that any copyright laws enacted can be readily enforced, and more involved types of international trade restraints are not further encouraged by the technology.

MEDLARS and Developing Countries

In developing countries, the situation regarding and needs for biomedical information are considerably different from those found in the United States and other developed nations. The issues surrounding NLM's provision of data bases and on-line services are indicative of a broader set of issues that are far more profound than the implications of gathering information technologies. These issues are rooted in the most basic problems of development, and cannot be addressed nor redressed simply through changes in information policies. They are selectively presented here as problems that must be solved if MEDLARS is to enjoy wider usage in developing countries, but they are barriers to the use of all information services and products.

Further, these issues are symptomatic of larger trends in the developed countries and their relations with the developing world. The developed countries, and the United States in particular, are moving in significant ways *toward* a social order, which is variously labeled as "postindustrial society," "communication era," "knowledge society," or "information society." In developing countries there is not yet a widespread awareness of these trends. But where the awareness exists the movement toward information society is viewed with some alarm. Namely, while developing countries were by and large moving after the model of developed countries into industrialization, the developed countries themselves started moving away from industrialization into an information-based society. The view from the developing world is that in this way the developed countries will continue to be "one up" on developing countries, no matter how successful industrialization may become. The hated dependence will continue.

The long-term trend is viewed as follows: in the past we (developing countries) have supplied the developed world with raw materials and cheap labor, while the developed world had industries and manufacturing as the basis of its prosperity. Now that we are gaining industries (and in the process depleting raw resources, selling our clean air and water, and endangering our environment) we may end up providing the developed world with relatively cheap industrial products and continuous cheap labor, while the developed world will have the control of information and communication technologies, and continue to build its prosperity. Information policy is becoming grounds for continuation and even expansion of the conflicts between the North and the South. Policies and actions on biomedical data bases, including MEDLARS, should be such to minimize this conflict.

Biomedical Data Bases: A Contribution to Development

Many countries consider data bases as part of foreign policy, even if the United States may not want to do so. Two aspects of foreign policy should be considered in relation to biomedical data bases and developing countries; the potential U.S. contribution to the process of development and the potential gain in biomedical information from developing countries.

The contribution to the process of development is twofold. Biomedical data bases have information that directly impinges on and potentially contributes to the health of people; national health efforts are a high priority in most countries and biomedical data bases can contribute to these efforts; and the U.S. data bases are of benefit to the world as a whole. The use of biomedical data bases contributes to an increase in the general skill and level of use of modern information technology and resources in a country, i.e., biomedical data bases can generally contribute to the attainment of the symbols of the age and benefit developing countries through every practical use of modern information technologies.

The gain to the United States is also twofold. Increasingly, developing countries themselves are producing valuable biomedical knowledge of universal utility to health, including health and biomedical research and development in the United States. However, this information is not widely distributed through international channels; thus, through cooperation in data bases, the United States can gain information that otherwise might not be obtained directly and easily. Cooperation in biomedical data bases is one step that contributes to the free flow of information on an international level and to a freer society. Thus, one of the general U.S. foreign policy interests can be positively affected through biomedical data bases.

There is considerable isolation in biomedical communication in developing countries which impedes the free flow of information. For instance, on the global level, English-speaking and U.S. scientific publications are by far predominant over all other languages and countries. They are also the most cited. It is not surprising then that scientists from developing countries try to place their best publications in well-known American, British, or other European journals rather than in their own national journals. This creates what is known as a "manuscript drain" (related to the well-known "brain drain"). Although native journals are very important for scientific communication within their realm, they are not prospering domestically and have an even harder time being recognized internationally (e.g., being covered by *Current Contents, Science Citation Index,* or *Index Medicus*) incorporation into these sources is highly coveted, bringing with it domestic prestige.

Bilateral Agreements and Developing Countries

As noted earlier, NLM signed its first bilateral agreement in the United Kingdom and Sweden in the mid-1960's. Also in the mid-1960's, through cooperation with the Pan American Health Organization (PAHO), the South American Regional Library of Medicine (BIREME) was established in Sao Paulo, Brazil, to serve Latin America. In the late 1970's, BIREME started its own computerized service from MEDLARS tapes for Brazil and South America. The success of cooperation with PAHO and BIREME had a lasting effect on the NLM perceptions of what may be ideal in cooperation between NLM and developing countries.

In contrast to PAHO, the agreements for services to developing countries with the World Health Organization (WHO) headquartered in Geneva, Switzerland, varied over the years. Various experimental operations were established, but they changed and lapsed over time. In various ways NLM has either provided WHO with direct access to MEDLINE (and other data bases) or conducted searches itself on behalf of WHO; however, no permanent mode exists for U.S. bibliographic data base services through WHO: no model for cooperation has evolved from those rather disappointing collaborations.

Presently, there are 13 countries in the world and one international organization, PAHO, with which NLM has bilateral agreements regarding U.S. data bases or MEDLINE access. In addition to PAHO-BIREME, which is oriented toward developing countries, 3 of those 13 countries are developing nations: Mexico, Colombia, and Kuwait. The agreements with Colombia and Kuwait were signed in 1981: Kuwait's is not yet completely operational. Thus, BIREME and Mexico are representative of developing countries in the bilateral agreements.

However, there is a second tier effect of MEDLARS and other data base services to developing countries over which NLM has no control nor is directly involved. Namely, some developed countries that have a bilateral agreement to mount on their own MEDLARS tapes *do* provide services to developing countries. For instance, Japan offers searching to Southeast Asian countries, Sweden to Poland, Germany to Yugoslavia, and Australia to Western Pacific countries. This trend is spreading, and soon a good number of developing countries may have MEDLARS (and other U.S. data base) access without direct agreements with NLM (but based on NLM agreements with some other country). The mode of arrangements for this second tier access varies from country to country providing the access: most of the time it is based on outright assistance.

Barriers to Effective Transfer and Use of MEDLARS in Developing Countries

If we accept that the basic aim of all information activities (including those connected with biomedical data bases) is for information to be utilized, then we have to consider that the minimum conditions for effective transfer and use of information for socioeconomic development include: a propensity on the part of decisionmakers and problem solvers (in the case of biomedical information: health workers and policymakers) to use information, which is based among other things, on recognition of the value of the need for information; a level of infrastructure (indigenous information systems, including libraries) that makes the right information first available and then accessible for use, and a type of political and social conditions which is favorable for encouraging the use of information and development of an information infrastructure.

In contrast to the United States and other industrialized countries, developing countries meet these conditions only partially and to highly variable degrees.

RECOGNITION

While European countries recognized the importance of biomedical data bases relatively early, and not long after initiation of MEDLARS began providing some form of MEDLARS access, widespread recognition is not evident in the developing countries. Recognition exists in international bodies and at international meetings, but in most developing countries themselves, the recognition is limited to a very few senior librarians and government officials. All working librarians have heard about MEDLARS and MEDLINE, but for most it is very far removed from their daily experience and problems.

Where this recognition exists, NLM and MEDLARS are held in absolutely highest regard. NLM has a great reputation in developing countries, and is perceived as the world's biomedical information resource. In no other subject area has the United States (or for that matter any other country) a single library and information institute that effects such a tremendous influence on the whole world.

As a result of this recognition, the expectations for NLM are also very high. At times these expectations are unrealistic. There is little or no undertanding of the U.S. governmental mechanism under which NLM operates, its responsibility to Congress and its domestic role, trials, and tribulations. The view that NLM is a world institution precludes seeing it as a U.S. domestic institution.

In turn, there is little recognition in the United States of NLM's tremendous influence, though unintended, on the world's biomedical information transfer in general, and on every country's medical libraries in particular. NLM's concern for developing countries has been defined through the looking glass of international organizations and the eyes of the world's professional committee or conference members, rather than by actual domestic situations and practicing librarians.

INFORMATION NEEDS AND COVERAGE

MEDLARS is primarily oriented toward the information needs of U.S. biomedical communities in research, clinical medicine, and health practice. Where similar communities are found in developing countries, their information needs are well covered by MEDLARS, particularly for researchers and clinicians. However, in many respects MEDLARS coverage is not fully adequate for the information needs of developing countries. For instance, MEDLARS covers only a fraction of biomedical literature published in developing countries, and thus can offer the world's biomedical information for use in developing countries, but not domestically published information. Eight countries contribute approximately 80 percent of the items in MEDLARS: over one-third of them are from the United States. Thus, MEDLARS does not and cannot serve as a domestic data base in the developing countries.

Hence, MEDLARS is not a data base that is oriented in a comprehensive way toward the needs of developing countries, nor is it so mandated. It can satisfy some of those perceived needs only to a degree. In this lies one of the problems of an indiscriminate use of MEDLARS in or for developing countries.

The success of BIREME, the South American Regional Medical Library organized by PAHO, in the publishing of *Latin American Index Medicus,* with a view toward developing it as a regional data base, has also brought to the surface expression of needs to publish similarly an *African Index Medicus,* and an *Index Medicus* in other world regions. These present rich opportunities for developing strong relations between NLM and the respective regions, relations similar to those that exist between NLM and Latin America.

A further problem in meeting the expressed information needs of developing countries stems from the fact that MEDLARS is a bibliographic data base. These needs might better be met by factual data bases, i.e., those providing actual information or data found in the literature and not simply bibliographies of literature containing data of information. The NLM Hepatitis Knowledge Base is a splendid example of such a data base. The production of such source data bases may possibly be a proper way for the private sector of the information industry or WHO to become involved in services for developing countries.

SELECTION PROBLEMS

MEDLARS does not cover all of the world's biomedical literature, but selects, in a comprehensive manner, biomedical literature deemed most substantive. However, in this selection no distinction is made among papers as to their relative quality, value, validity, usefulness, redundancy, etc. A Nobel prize-winning paper receives the same attention as one with no or questionable value. Selection is based only on a given article's relevance to a particular subject or its source journal, and gives MEDLARS an important characteristic—comprehensiveness.

However, developing countries have expressed desires for data bases with selectivity-based quality, appropriateness, value, and similar attributes of information. MEDLARS cannot satisfy such desires, nor should it be expected to do so. But, it may be possible to create out of MEDLARS smaller data bases satisfying these different criteria of selectivity. Again, it is an open question as to who should create such selective sub-data bases. This may be done at NLM but may also be accomplished by the private sector, developing countries themselves, international organizations, or consortia.

At the root of the problem is the fact that MEDLARS searches, and searches in other bibliographic systems, often retrieve too much information for an individual to deal with and absorb. Although MEDLARS capabilities allow the user to tailor a search so that a broad search which retrieves many citations or a narrower, more focused retrieval can be obtained, many factors affect the character and quality of the search performed. These include the precision of the search question, the amount of information on the topic of the search, the quality of the indexing The size of MEDLARS is large and growing, and so is the output. Thus, ultimately selectivity of the type described is aimed at producing outputs that on the one hand are within the realm of human scale for dealing with information (offering options by which an individual can adjust the threshold for that scale) and on the other hand are reasonably the best (or at least better) representatives of what is known on the topic requested.

ACQUISITION PROBLEMS

One of the elements of NLM's policy is to have a central organization in each country with a bilateral agreement which will undertake all responsibility for mounting of and services from MEDLARS tapes or access to MEDLINE; this organization must meet various technical criteria, as enumerated. Very few, if any developing countries can meet these; thus, this requirement effectively precludes developing countries from entering into bilateral agreements as presently instituted.

Most developing countries simply do not have an organization that is capable of meeting the technical, fiscal, and personnel requirements. Making arrangements with Ministries of Health as central organizations makes little or no sense, because they are not in the information business, and although they have influence and play a most important role in health and the dissemination of health-related information, they have little or no connection with local libraries and information users. Furthermore, many countries are not capable of meeting the quid pro quo requirements as instituted at present.

Even where a large medical library exists in a country or a region, agreement with such a library does not guarantee access to other libraries, because operational connections and cooperation among libraries is, in general, low. Also, most developing countries have inadequate collections for coping to any significant extent with document delivery problems. This presents a most frustrating aspect for users who find citations with electronic speed and then must wait months for documents to arrive from abroad by surface mail, if they ever arrive at all.

Thus, the present policy requirements work as a barrier, if not as an exclusion, for most developing countries to acquire some form of access to U.S. biomedical bibliographic data bases.

The regional approach to acquiring MEDLARS access, as in the case of BIREME, certainly has great merits, but also has great limitations. Regional structure and cooperation in Latin America is not at all the same as in the United States. Thus, besides the expected technical and personnel problems, BIREME has a great problem of finding an effective mode for service to the libraries and users in the region.

So far, the private sector of the information industry has used the same policies for acquiring access to their biomedical data bases (including MEDLINE) as in the United States. No specific policies for developing countries has been implemented or even debated. Private industry as yet seems to be unprepared to approach, in a significant way, the potential information markets in developing countries.

In conclusion, no one approach to acquiring access to U.S. bibliographic data bases by developing countries seems to bear as much fruit as the policies worked out with developed countries. It seems that more flexibility and experimentaiton may be in order, if success in this area is to be realized.

ACCESSIBILITY PROBLEMS

Intellectual Organization.—Intellectual organization of U.S. bibliographic data bases based on MeSH (medical subject headings) seems to be satisfactory for developing countries, though it does not conform in many instances to local desires and needs. However, MEDLARS, with its standardized terminology, classification and high quality of application, and indexing seems to have an additional beneficial effect wherever introduced: it adds to standardization of intellectual organization. Resource-sharing for intellectual organization (cataloging, indexing) is low in developing countries. MEDLARS has not yet changed this in any significant way; however, the potential is there. It is possible that a wider application of U.S. bibliographic data bases may stimulate implementation of resource-sharing plans.

Another issue in this area are indexes. *Index Medicus,* produced by NLM, is the most widely accepted indexing tool in biomedicine in developing countries. However, it is widely recognized that *Index Medicus,* although fully adequate for general medical (particularly research and clinical) needs, does not fully cover local publications, needs, and language. Thus, national or regional indexing tools are urged with great frequency and vehemence. In 1979, after many years of discussion and 2 years of preparation, with great help of NLM, and under the auspices of PAHO, the *Latin American Index Medicus (IMLA)* started publication by the South American Regional Library of Medicine (BIREME) in Sao Paulo, Brazil. The event is considered with unabashed jubilation as a breakthrough for developing countries—and rightly so. It is also considered a model for what can be achieved through cooperation between all parties concerned. A similar *African Index Medicus* is being urged, as well as indexes for other regions.

Printed indexes and data bases are at present closely connected. *Index Medicus* whets the appetite for MEDLARS data bases, MEDLINE in particular. *Latin American Index Medicus* is a prelude for a regional data base.

Physical Access.—Mexico is the only developing country that has up to the present implemented physical access to MEDLINE through NLM computers and BIREME. In turn Mexico provides further domestic access through a central institution in a way which has a restrictive effect on the number of libraries and other institutions accessing MEDLINE. BIREME has limited on-line capacity: there are terminals at four subcenters —at BIREME itself and three other cities in Brazil. Online access from other Latin American countries was and is still being planned, but telecommunication and other difficulties have as yet precluded the implementation.

In many cases, the access from developing countries to MEDLARS centers around the world (including accessing NLM for MEDLINE searching) is by international mail. All the implications of slow and bad mail service pertain to this mode of access. Turnaround time is long, and interaction nonexistent. Mail access to computerized searching has not proved effective or desirable, but it is still better than no access. If well organized, it can be a stop-gap measure.

Finally, the physical access to comprehensive library collections to satisfy document delivery is also very difficult for the majority of present or potential online users in developing countries. It also depends on mail. The turnaround time for requests to be fulfilled is often 2 months, and a goodly number are not filled at all or lost in shipment.

Dissemination and Service.—Libraries in developing countries generally offer a limited number of services. On-line searching, even where possible directly or indirectly by mail, is not yet a service that has penetrated libraries to any significant extent. NLM has relatively little dealings with the mass of libraries in developing countries; thus, it cannot directly affect use in such services.

Information Technology.—Information technology presents a formidable obstacle to further proliferation of MEDLARS centers in developing countries. It is not so much the lack of computers, but the lack of capacity to make the technology work. Transfer of ELHILL software for running NLM's data bases is wrought with great difficulties, however, it is not ELHILL alone that poses difficulty—all similar software transfers for all kinds of information retrieval system have similar problems. Closely connected is the problem of technologically competent personnel. No solution has been found by NLM (or for that matter by any institution involved in information retrieval) for more effective and faster transfer of technology, software, and hardware, for establishment of MEDLARS centers in developing countries. Many calls for simplified software have not been answered.

The technological problems for direct on-line access to MEDLINE from developing countries is centered around telecommunications. Where TELENET or TYMNET nodes exist this is not that difficult. However, where they do not exist it is impractical, if not impossible, to establish effective direct access for reasons of high costs, restructive regulations, and poor technical quality and reliability.

Information Networks.—As yet, the NLM network (including NLM connected networks such as BIREME or those in Europe and Japan) is the only true international network affecting developing countries at least to some extent in the realm of any and all biomedical data bases. Private U.S. vendors that have mounted MEDLINE (and other biomedical data bases) began offering these data bases to developing countries through their network or networks. However, as yet they have not penetrated this market.

It is quite hard to consider NLM being in competition with the U.S. private information industry in developing countries, because the structure of medical libraries and all health institutions in developing countries is entirely government-affiliated and dependent, and is very much influenced in action and philosophy by WHO and other regional health organizations. As such, this very structure seeks governmental and not private connections, structure is the only one through which network connections can eventually be established. Thus, if NLM were eliminated from or restricted in data base cooperation with developing countries, there would not be any biomedical data base services in developing countries in the foreseeable future, and those meager beginnings made would be eliminated.

UTILIZATION

The use of MEDLARS centers or MEDLINE searching from developing countries is low, and in comparison with the total number of searches performed on MEDLARS the number is probably very small. For example, BIREME performs about 2,500 annual searches, some 70 percent from Brazil, the rest from countries throughout South America (78). The Southeast Asian Medical Information Center (under auspices of the International Medical Foundation of Japan) reports, for the period of August 1979 to September 1980, having filled 38 search requests coming from five countries in the region (93). However, at present a direct comparison between levels of use of U.S. data bases in developing countries with those in the United States is clearly not valid. What should be considered is the long-term potential and benefit, not present use.

Neither NLM nor any other institution has been involved with user education, marketing, or promotion in developing countries in respect to U.S. data bases. While a number of librarians in developing countries have been exposed to MEDLARS and MEDLINE (or at least heard of it), not many potential users have. This may also account for low present use.

Some other U.S. agencies, the National Technical Information Service (NTIS) in particular, have been engaged for some time in marketing and promoting their scientific and technical information products and services in developing countries to a much greater extent than NLM. NTIS created special bibliographies and newsletters, some translated into Spanish, special promotional material, traveling presentations and demonstrations and the like, for marketing and promotion in developing countries. Some similar or even joint approach may be desirable for U.S. biomedical data bases.

INFORMATION PROFESSIONALS

NLM has trained a number of librarians and information scientists in MEDLARS construction and MEDLINE searching This is one of the most highly praised and desirable activities. Requests for more training of personnel are made with increased frequency and urgency. NLM has developed a very good training program, probably better than that of any other U.S. information agency.

Still, as mentioned, the vast majority of biomedical librarians in developing countries are not knowledgeable of MEDLARS, particularly as to operational aspects—e.g., constructive searches, structure of vocabulary, etc. As extensive as NLM's training efforts are, they have penetrated only skin deep. It appears that MEDLARS foreign centers in developing countries or at BIREME have not yet conducted any extensive amounts of training on their own.

Final Remarks

Although most of the issues concerning the use of MEDLARS in developed and developing countries appear to vary widely between the two disparate spheres, they share strategic issues. The issues that transcend data base issues are: the evolution of U.S. society into an information society, information as a national symbol and element in foreign policy, free flow of information in general and in biomedicine in particular, transborder data flow and the restrictions imposed against that flow, and cooperation rather than competition as a base for information activities.

There are vast dissimilarities, however, between the information problems and the issues between developed and developing areas, and even among individual countries, and the activities of the United States and NLM in each of the areas. In general, MEDLARS is being used effectively for the public's health in developed countries. MEDLARS and other data bases have so far exercised a minimal operational impact on developing countries, but they have a high and still growing impact on expectations in developing countries. Thus, policy decisions should also be carefully weighted in relation to these expectations and needs.

More thought must be given to means of increasing MEDLARS' utilization in developing nations. It is beyond the scope of this report to consider the basic problems associated with development that lie at the root of the problem, but it should be noted that technological advances in computation may be an aid to improvement.

As noted previously, the technological problems for direct on-line access to MEDLARS data bases from developing countries is centered around telecommunication. This problem may be ameliorated by future trends in computerized data bases (see app. H). Among them are the widespread use of microcomputers and their potential low cost. At the same time there has been the development of the video disk, with which one can access most of a data base on a microcomputer without waiting for time-sharing connections or paying time charges. The disk would also be relatively inexpensive. Applying these technologies to developing countries may be one step in the major and complex problems in improving the transfer of biomedical information.

Appendix J.—Health Program Advisory Committee and Acknowledgments

---

HEALTH PROGRAM ADVISORY COMMITTEE

Sidney S. Lee, *Chairman*
Vice President, Michael Reese Hospital and Medical Center

Stuart H. Altman
Florence Heller School
Brandeis University

Robert Ball*
Institute of Medicine
National Academy of Sciences

Lewis N. Butler*
Health Policy Program
School of Medicine
University of California, San Francisco

Kurt Deuschle
Mount Sinai School of Medicine

Zita Fearon*
Consumer Commission on the Accreditation of Health Services, Inc.

Rashi Fein
Center for Community Health and Medical Care
Harvard Medical School

Melvin A. Glasser
Committee for National Health Insurance

Patricia King
Georgetown Law Center

Joyce C. Lashof
School of Public Health
University of California, Berkeley

Mark Lepper
Vice President for Inter-Institutional Affairs
Rush-Presbyterian-St. Luke's Medical Center

Margaret Mahoney
President
The Commonwealth Fund

Frederick Mosteller
Department of Health Policy and Management
Harvard School of Public Health

Beverlee Myers*
Director
Department of Health Services
State of California

Mitchell Rabkin
General Director
Beth Israel Hospital

Dorothy P. Rice
Department of Social and Behavioral Sciences
School of Nursing
University of California, San Francisco

Richard K. Riegelman
School of Medicine
George Washington University

Walter L. Robb
Vice President and General Manager
Medical Systems Operations
General Electric

Frederick C. Robbins
President
Institute of Medicine
National Academy of Sciences

Rosemary Stevens
Department of History and Sociology of Science
University of Pennsylvania

Kerr White
Deputy Director for Health Sciences
Rockefeller Foundation

---

*Former members of the Health Program Advisory Committee whose terms expired prior to the completion of this study.

ACKNOWLEDGMENTS

This technical memorandum was prepared with the advice and review of not only the advisory panel for the study but the Health Program Advisory Committee and many other people. The Director of the National Library of Medicine, Martin M. Cummings, the Deputy Director, Kent Smith, and numerous other members of the NLM staff were extremely helpful. OTA staff would also like to thank the following individuals for their valuable contributions.

Scott Adams
Louisville, Ky.
Ken Allen
Office of Management and Budget
William Baker
Bell Telephone Laboratories
Howard Bleich
Beth Israel Hospital
Jane Bortnick
Congressional Research Service
Yale Braunstein
Brandeis University
John Bruer
Macy Foundation
Michael D. Cooper
University of California
Melvin Day
National Technical Information Service
Nancy DeWath
University of California
Jan Egeland
Bibliographic Retrieval Services
Nancy Fazzone
Salem Hospital
Gail Fisher
National Center for Health Statistics
Dorothy Fitzgerald
University of Western Ontario
Allan Fox
Kaye, Scholar, Fierman, Hays & Handler
Oswald Ganley
Harvard University
Carole Ganz
National Science Foundation
James Hahn
Veterans Administration
Joseph Heinmiller
National Institute of Education
Sherrill Hites
Harvard Medical School
Mary Horres
University of North Carolina
David Hoyt
National Agricultural Library
Edward J. Huth
American College of Physicians
Laura Kassebaum
Bibliographic Retrieval Services
Gertrude Lamb
Hartford Hospital
Joan Marshall
McMaster University
Nina Matheson
Association of American Medical Colleges
Grace McCarn
H. W. Wilson Co.
H. Calvin Meadows
National Health Planning Information Center
Peggy Miller
Kaye, Scholer, Fierman, Hays and Handler
Lynne Neufeld
Federation of Abstracting & Indexing Services
L. Daniel O'Neill
Advanced Computer Techniques
Raymond Palmer
Medical Library Association
Michael Pollard
Federal Trade Commission
Johnathan Ruby
Elscint Corp.
William Schecterley
General Accounting Office
Charles W. Sergeant
Texas Technical University
Winifred Sewell
University of Maryland
Vladimir Slamecka
Georgia Institute of Technology
John Timour
University of Pennsylvania Medical School
Richard Tuey
Department of Health and Human Services
James Ullom
National Center for Health Services Research
Rick Weingarten
Office of Technology Assessment
William J. Welsh
The Library of Congress
Jane Sisk Willems
Office of Technology Assessment
Martha Williams
University of Illinois
Richard Willard
Information Industry Association
Marvin Wilson
National Technical Information Service
Paul Zurkowski
Information Industry Association

Appendix K.—Acronyms, Abbreviations, and Glossary

Acronyms and Abbreviations

AAMC	Association of American Medical Colleges
AGRICOLA	Agricultural On-Line Access (NAL)
AIM-TWX	Abridged Index Medicus via the Teletypewriter Exchange Network
AMA	American Medical Association
APHA	American Public Health Association
BCN	Biomedical Communications Network
BRS	Bibliographic Retrieval Services
CAS	Chemical Abstracts Service
CIJE	*Current Index to Journals in Education*
CIM	*Cumulated Index Medicus*
CITE	Current Information Transfer in English
COSATI	Committee on Scientific and Technical Information
DHEW	Department of Health, Education, and Welfare (now DHHS)
DHHS	Department of Health and Human Services
DIALOG*	DIALOG Information Services, Inc.
ERIC	Educational Resources Information Center (NIE)
FAMLI	*Family Medicine Literature Index*
GAO	General Accounting Office (U.S. Congress)
GPO	Government Printing Office
HEW	see DHEW
HHS	see DHHS
IRL	Information Retrieval, Ltd.
ISDN	integrated services digital network
ISI	Institute for Scientific Information
MEDLARS	Medical Literature Analysis and Retrieval System
MEDLINE	MEDLARS On-Line
MeSH	medical subject headings
MLAA	Medical Library Assistance Act
NAL	National Agricultural Library (USDA)
NAS	National Academy of Sciences
NCLIS	National Commission on Libraries and Information Science
NHPIC	National Health Planning Information Center
NIE	National Institute of Education (Department of Education)
NIH	National Institutes of Health (DHHS)
NLM	National Library of Medicine (NIH)
NMAC	National Medical Audiovisual Center (NLM)
NTIS*	National Technical Information Service (Department of Commerce)
OECD	Organization for Economic Corporation and Development
OMB	Office of Management and Budget
OTA	Office of Technology Assessment (U.S. Congress)
PAHO	Pan American Health Organization
RIE	Research in Education
RML	Regional Medical Library
RMLP	Regional Medical Library Program
SATCOM	Scientifc and Technical Communication
SDC	System Development Corp.
SDI	selective dissemination of information
SUNY	State University of New York
USDA	U.S. Department of Agriculture
WHO	World Health Organization

Glossary

Batch searching: A method of information processing whereby data are read into a computer from keypunched cards. The desired program, usually stored on magnetic tape, is entered from the tape into the computer memory. The program acts on the data, and the results are made available on a printer or a second set of keypunch cards. Unlike on-line searching, the batch method does not allow the user to interact directly with the computer to change commands while the search is being done.

Boolean logic: A system for expressing relationships between concepts using the connectives "and," "or," and "not." Many information retrieval systems use Boolean logic as a method of searching data bases.

Cataloging: The process of describing and classifying books or other library materials.

Data base: An organized collection of information, usually on a specific subject, in machine-readable form and accessible by computer.

Distribution computer system: An arrangement of computers in which the computer complex has many separate computing facilities all working in a cooperative manner, rather than a single computer in a single location. The system is versatile, and small computers in geographically dispersed locations can be used for simple tasks in conjunction with a powerful large computer that is used for larger tasks.

End users: Individuals, such as physicians and researchers, who request and use on-line search results.

---

*This is not only the short name of the organization specified but is also the name of a data base.

Indexing: The process of assigning headings (from a specialized thesaurus) to articles that have been analyzed. The key words assigned are then used in the retrieval process to identify the relevant citations in a bibliographic search.

Intermediate users: Librarians and other information specialists trained to conduct on-line searches of computerized data bases.

MEDLARS: NLM's computerized retrieval and technical processing system, a complex IBM multiprocessing system that maintains data files, provides on-line retrieval services, and produces computer photocomposed publications. The current system is often referred to as MEDLARS II; the system under development is MEDLARS III.

MEDLINE: The original, largest and most utilized MEDLARS data base. MEDLINE contains references to biomedical and other literature relevant to health and health services.

Natural language: Standard English.

On-line: A computer program is considered to be "on-line" if it can respond to commands while the information specialist is working with the program at the computer terminal. "On-line" capability allows for near-instantaneous response, and thus allows the user to interact with the computer.

Periodical: A publication, such as a journal, which is issued at fixed intervals usually longer than a day between issues or numbers.

Precision ratio: The number of relevant items retrieved in a bibliographic search divided by the total number retrieved in the search.

Recall ratio: The number of relevant items retrieved in a bibliographic search divided by the number of relevant items indexed by the system.

Regional Medical Library: A library that provides document delivery services and assistance in organizing collections to smaller hospital and medical school libraries in a given geographic region under a contract awarded by NLM. Currently, the United States is divided into 11 regions, but NLM plans to reorganize the Regional Medical Library Program into 7 by fall 1982.

Serial: A publication issued as one of a consecutively numbered and indefinitely continued series. It may be a periodical or a publication, such as a monograph or conference proceeding, that is issued at irregular intervals.

Terminal: A machine similar to a typewriter which is capable of transmitting and receiving electronic signals to/from a computer.

Text word searching: The process of searching a data base using words that appear in the title, abstract, or text of the article cited. Text word searching does not require that an article be indexed using terms from a subject heading list, i.e., a controlled vocabulary.

Two-tiered searching: The process of searching a data base using both a controlled vocabulary (used for indexing articles) and words appearing in the title or abstract or the text of the article cited.

References

References

1. Abelson, P. H., "Electronics and Scientific Communication," *Science* 17(210):255, 1980.
2. Ackoff, R. L., Cowan, T. A., Davis, P., et al., *Designing a National Scientific and Technological Communication System* (the SCATT Report) (Philadelphia: University of Pennsylvania Press, 1976).
3. Adams, S., former Deputy Director, National Library of Medicine, personal communication, February 1981.
4. Adams, S., *Medical Bibliography in an Age of Discontinuity* (Chicago: Medical Library Association, Inc., 1981).
5. Adler, A., "Excerpta Medica," statement before the Senate Committee on Labor and Human Resources, Apr. 1, 1982.
6. Algermissen, V., "Biomedical Librarians in a Patient Care Setting at the University of Missouri—Kansas City School of Medicine," *Bull. Med. Libr. Assoc.* 62:354, 1974.
7. Allen, K., Office of Management and Budget, personal communication, November 1981.
8. American Library Association Council, "Resolution on Continued Free Public Access to Federal Information," Council Document No. 71.2, adopted in San Francisco, Calif., July 1, 1981.
9. American Psychological Association, statement on H.R. 4758, presented to the Subcommittee on Government Information and Individual Rights, House Committee on Government Operations, Feb. 25, 1982.
10. Antes, J., "The Rural Area Hospital Can Afford a Librarian," *Bull. Med. Libr. Assoc.* 70:233, 1982.
11. Bachrach, C., and Charen, T., "Selection of MEDLINE Contents, the Development of Its Thesaurus, and the Indexing Process," *Med. Infor.* 3(3):237, 1978.
12. Baker, C., King, S., and Wanger, J., "Human Services Information Clearinghouses: A Discussion of Policy Issues," contract No. HEW-100-79-0183, prepared for the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, September 1981.
13. Baker, W. O., Bell Telephone Laboratories, Murray Hill, N.J., personal communication, January 1981.
14. Beckelhimer, M. A., Cox, J. W., Hutchins, J. W., et al., "The MEDLINE Hardware and Software," *Med. Inf.* 3:197, 1978.
15. Becker, J., *A National Approach to Scientific and Technical Information in the United States* (Los Angeles: Becker & Hayes Publishing Co., 1976).
16. Beringer, D., and Adkinson, B., "Interaction Between the Public and Private Sectors in National Information Programs," *Ann. Rev. Inf. Sci. & Tech.* 13:3, 1978.
17. Bernstein, L., Siegel, E., and Goldstein, C., "The Hepatitis Knowledge Base. A Prototype Information Transfer System," *Ann. Int. Med.* Part 2:165, July 1980.
18. Bourne, C., "On-Line Systems: History, Technology, and Economics," *J. Am. Soc. Inf. Sci.* 31(3):155, 1980.
19. Bourne, C., *Technology in Support of Library Science and Information Service* (Bangalore, India: Sarada Ranganathan Endowment for Library Science, 1980).
20. Brandt, E., Jr., "Health Services: A Shared Responsibility, Reconsideration of the Federal Role," presented at the 1981 Annual Meeting of the Institute of Medicine, National Academy of Sciences, Washington, D.C., Oct. 28, 1981.
21. Braunstein, Y., "The Functioning of Information Markets," in *Issues in Information Policy,* J. Yurow, project director (Washington, D.C.: Department of Commerce, February 1981).
22. Brown, C. P., "On-Line Bibliographic Retrieval Systems Use," *Spec. Libr.* 68(4):155, 1977.
23. Brown, G. E., Jr., "Information and Society: A Congressional View," remarks before the American Society for Information Science, Washington, D.C., Oct. 27, 1981.
24. Bruer, J., "Subject Coverage and Recall Performance of MEDLARS for Health Sciences and Services," contract report submitted to the Office of Technology Assessment, U.S. Congress, 1981.
25. Burrows, S., and Kyle, S., "Searching the MEDLARS File on NLM and BRS: A Comparative Study," *Bull. Med. Libr. Assoc.* 67(1):15, 1979.
26. Calkins, M. L., "On-Line Services and Operational Costs," *Spec. Libr.,* January 1977.
27. Cheshier, R., Director, Cleveland Health Sciences Library, personal communication, June 3, 1981.
28. Cooper, M., School of Library and Information Studies, University of California, Berkeley, personal communication, April 1981.
29. Cooper, M., "Charging Users for Library Service," *Inf. Proc. & Mgt.* 14:419, 1978.
30. *Coping With the Biomedical Literature Explosion: A Qualitative Approach,* K. Warren and W. Goffman (eds.) (New York: Rockefeller Foundation, 1978).
31. Corning, M. E., "International Biomedical Communications: The Role of the United States Na-

tional Library of Medicine," *Health Commun. Info.* 6:212, 1980.

32. Corning, M. E., and Cummings, M. M., "Biomedical Communications," in *Advances in American Medicine: Essays at the Bicentennial,* J. Z. Bowers and E. Purcell (eds.) (New York: Josiah Macy, Jr., Foundation, 1976).
33. Cummings, M., Director, National Library of Medicine, personal communication, February 1982.
34. Cummings, M., "Editorial," *Med. Inf.* 3:164, 1978.
35. Cummings, M., "The National Library of Medicine and Quality Information Filters," in *Coping With the Biomedical Literature Explosion,* K. Warren and W. Goffman (eds.) (New York: Rockefeller Foundation, 1978).
36. Cummings, M., "The U.S. National Library of Medicine: A National Resource Serving International Needs," in *Selective Libraries for Medical Schools in Less-Developed Countries,* T. Saracevic (ed.) (New York: Rockefeller Foundation, 1980).
37. Cummings, M., and Corning, M. E., "The Medical Library Assistance Act: An Analysis of the NLM Extramural Programs, 1965-1970," *Bull. Med. Libr. Assoc.* 59:375, 1971.
38. Department of Commerce, *Issues in Information Policy,* J. Yurow, project director (Washington, D.C.: Department of Commerce, February 1981).
39. Dersimonian, R., Charette, J., McPeek, B., et al., "Statistical Information in Comparative Clinical Trials," unpublished, reported by F. Mosteller in the foreword to the *Milbank Mem. Fund Q./Health & Soc.* 50(3), summer 1981.
40. DeWath, N., *Demand for Public Library Services: A Time Allocation and Public Finance Approach to User Fees,* doctoral thesis, University of California, Berkeley (Ann Arbor, Mich.: University Microfilms International, 1980).
41. Doszkocs, T., Rapp, B., and Schoolman, M., "Automated Information Retrieval in Science and Technology," *Science* 208:25, 1980.
42. Dresner, Morris & Tortorello Research, *The Business of Information Report* (Washington, D.C.: Information Industry Association, 1980).
43. Dunn, D., Professor of Engineering Economic Systems, Stanford University, memorandum, to the Advisory Panel, MEDLARS and Health Information Policy, Office of Technology Assessment, U.S. Congress, Jan. 8, 1982.
44. Eastwood, S., Department of Neurological Surgery, University of California, San Francisco, "The Other Gatekeeper," unpublished, 1981.
45. Egeberg, R. O., Assistant Secretary for Health and Scientific Affairs, "Sale of Magnetic Tapes From the MEDLARS System to Profitmaking Institutions," memorandum to the Secretary of the Department of Health, Education and Welfare, Dec. 10, 1970.
46. Egeland, J., President, Bibliographic Retrieval Services, Latham, N.Y., personal communication, May 1982.
47. Etzioni, A., "The Need for Quality Filters in Information Systems," *Science* 171:133, 1971.
48. Executive Office of the President, Bureau of the Budget, *Circular A-25,* "User Changes," Sept. 23, 1959.
49. Executive Office of the President, Office of Management and Budget, *Circular A-76* revised, "Policies for Acquiring Commercial or Industrial Products and Services Needed by the Government," 1979.
50. Fazzone, N., Director, Library Services, Salem Hospital, Salem, Mass., personal communication, January 1982.
51. Fink, D. A., "An Evaluation of the National Library of Medicine (NLM) Training Grant Program," contract No. HUMMRRO-CR-01-73-5, prepared by the Human Researches Research Organization for the National Library of Medicine, September 1973.
52. Fitzgerald, D., "Development of an International Index in Family Medicine," presented at the Conference of the Medical Library Association, Montreal, Canada, June 1982.
53. Freeman, D., "Interaction Marketing," *Ann. Rev. Inf. Sci. & Tech.* 13:36, 1978.
54. Garfield, E., "Current Comments: Excerpta Medica—Abstracting the Biomedical Literature for the Medical Specialist," *Current Contents* (28):5, 1980.
55. General Accounting Office, U.S. Congress, *Better Information Management Policies Needed: A Study of Scientific and Technical Bibliographic Services* (Washington, D.C.: U.S. Government Printing Office, Aug. 6, 1979).
56. General Accounting Office, Human Resources Division, U.S. Congress, "National Library of Medicine's Medical Literature Analysis and Retrieval System" (Washington, D.C.: U.S. Government Printing Office, Apr. 15, 1982).
57. Goffman, W., Bruer, J., and Warren, K., *Research on Selective Information Systems: A Bellagio Conference, October 23-27, 1979* (New York: Rockefeller Foundation, September 1980).
58. Greenberg, B., Breedlove, R., and Berger, W., "MEDLINE Demand Profiles: An Analysis of Re-

quests for Clinical and Research Information," *Bull. Med. Libr. Assoc.* 65(1):22, 1977.

59. Greenberger, M., et al., "Making Technical Information More Useful: The Management of a Vital National Resource," report to the Chairman of the Federal Council for Science and Technology, Washington, D.C., 1972.
60. Guiliano, V., Ernst, M., Dunlop, J., et al., *Into the Information Age: A Perspective on Federal Action on Information* (Chicago: American Library Association, 1978).
61. Gunning, J. E., Fierberg, J., Goodchild, E., et al., "Use of an Information Retrieval Service in an Obstetrics/Gynecology Residency Program," *J. Med. Educ.* 55:120, 1980.
62. Harris, D. K., personal communication, September 1981.
63. Hitchingham, E. E., "MEDLINE Use in a University Without a School of Medicine," *Spec. Libr.* 67(4):188, 1976.
64. Horowitz, G., and Bleich, H. L., "Paperchase: A Computer Program To Search the Medical Literature," *N. Eng. J. Med.* 305:925, 1981.
65. Hubbard, W. N., President, The Upjohn Co., Kalamazoo, Mich., personal communication, Nov. 11, 1981.
66. Information Industry Association, *The Business of Information Report,* prepared by Dressner, Morris & Tortorello Research (Washington, D.C.: Information Industry Association, 1980).
67. Ingelfinger, F., "Peer Review in Biomedical Publication," *Am. J. Med.* 56:686, 1974.
68. Kassebaum, L., and Leiter, J., "Training and Continuing Education for On-Line Searching," *Med. Infor.* 3(3):165, 1978.
69. Kennedy, E., President of BIOSIS, Philadelphia, personal communication, Jan. 8, 1982.
70. Kenton, C., and Scott, Y., "MEDLINE Searching and Retrieval," *Med. Infor.* 3(3):225, 1978.
71. King, D. W., and McDonald, D., "Federal and Non-Federal Relationships in Providing Scientific and Technical Information," prepared for the National Science Foundation by King Research, Rockville, Md., March 1980.
72. King, D. W., and Roderer, N., *A Study of Pricing Policies for Information Products and Services,* prepared for the United States Geological Survey, National Cartographic Information Center by King Research, Rockville, Md., March 1978.
73. Kozmetsky, G., and Ruefli, T. W., *Today's Leadership Initiatives: Decisions That Cannot Wait* (New York: The Conference Board, 1971).
74. Kunz, J., "Index Medicus: A Century of Medical Citation," *J.A.M.A.* 241(4):387, 1979.
75. Lamb, G., former President, Medical Library Association, personal communication, November 1981.
76. Lancaster, F. W., Professor of Library Science, University of Illinois, personal communication, September 1981.
77. Lancaster, F. W., *Evaluation of the MEDLARS Demand Search Service* (Bethesda, Md.: National Library of Medicine, 1968).
78. Lancaster, F. W., "The Evaluation of Published Indexes and Abstract Journals," *Bull. Med. Lib. Assoc.* 59:479, 1971.
79. Lancaster, F. W., *Information Retrieval Systems: Characteristic Testing and Evaluation* (New York: John Wiley & Sons, 1968).
80. Lave, L. B., and Fedorowicz, J., *Government Policy Toward the Dissemination of Scientific and Technical Information* (Pittsburgh: Graduate School of Industrial Administration, Carnegie-Mellon University, April 1980).
81. Ling, S., *An Econometric Analysis of NLM On-Line Services,* prepared for National Library of Medicine by Teleconsult, Inc., Washington, D.C., September 1975.
82. Maina, W. E., Jenkins, C. G., and Meakin, F. A., "The CLR/NLM Health Sciences Library Management Intern Program: First Year," *Bull. Med. Libr. Assoc.* 68:16, 1980.
83. Marshall, J., McMaster University, personal communication, May 1982.
84. Matheson, N. W., and West, R. T., "NLM Medical Library Resource Improvement Grant Program: An Evaluation," *Bull. Med. Libr. Assoc.* 64:309,1976.
85. McCarn, D., "MEDLINE Users, Usage, and Economics," *Med. Infor.* 3(3):177, 1978.
86. McCarn, D., "National Library of Medicine—MEDLARS and MEDLINE," in *Encyclopedia of Computer Science and Technology,* J. Belzer, A. Holzman, and A. Kent (eds.) (New York: Marcel Dekker, Inc., 1978).
87. McCarn, G., "The On-Line User Network: Organization and Working Procedures," *Med. Infor.* 3(3):211, 1978.
88. McCarthy, S. E., Maccabee, S. S., and Feng, C. C., "Evaluation of MEDLINE Service by User Survey," *Bull. Med. Libr. Assoc.* 62:367, 1974.
89. McGee, J. L., "A MEDLINE Feasibility Study," *Bull. Med. Libr. Assoc.* 68:278, 1980.
90. Meltsher, A. J., "Local Revenue: A Political Problem," in *Financing the Metropolis Public Policy*

in Urban Economics, J. P. Crecines (ed.) (Beverly Hills, Calif.: Sage Publications, 1970).

91. Mid-Atlantic Regional Medical Library Program, National Library of Medicine, *Newsletter,* November 1981.
92. Milliman, J. W., "Beneficiary Charges—Toward a Unified Theory," in *Public Prices for Public Products,* S. Mushkin (ed.) (Washington, D.C.: Urban Institute, 1972).
93. Moll, W., "MEDLINE Evaluation Study," *Bull. Med. Libr. Assoc.* 62:1, 1974.
94. "NLM MEDLINE Wrestles With Anti-Discrimination," *Monitor* (8):1, October 1981.
95. National Academy of Sciences, Committee on Scientific and Technical Communication, *Scientific and Technical Communication: A Pressing National Problem and Recommendations for Its Solution* (the SATCOM Report) (Washington, D.C.: National Academy of Sciences, 1969).
96. National Commission on Libraries and Information Science, *Public Sector/Private Sector Interaction in Providing Information Services,* report of the Task Force on Public Sector/Private Sector Interaction (Washington, D.C.: U.S. Government Printing Office, February 1981).
97. National Commission on Libraries and Information Science, *Towards a National Program for Library and Information Services: Goals for Action* (Washington, D.C.: U.S. Government Printing Office, May 1975).
98. National Library of Medicine, letter to the Office of Technology Assessment, U.S. Congress, Apr. 12, 1982.
99. National Library of Medicine, "Implementation of OMB Circular JA-25 User Charges," letter from K. Smith, Assistant Director for Administration, NLM, to the Associate Director for Administration, NIH, May 16, 1978.
100. National Library of Medicine, Board of Regents Meeting, Bethesda, Md., May 28-29, 1981.
101. National Library of Medicine, Board of Regents Meeting, Bethesda, Md., Oct. 30, 1981.
102. National Library of Medicine, Board of Regents Meeting, Bethesda, Md., Feb. 4-5, 1982.
103. National Library of Medicine, "Interim Report of the NLM Task Force on Full Cost Recovery," Aug. 14, 1981.
104. National Library of Medicine, MEDLARS III Task Force, "MEDLARS III Overview," Sept. 30, 1980.
105. National Library of Medicine, "NLM Position on Office of Technology Assessment Study," letter to the Office of Technology Assessment, U.S. Congress, Apr. 29, 1981.
106. National Library of Medicine, "National Library of Medicine Regional Library Program Policy Statement," *Bull. Med. Libr. Assoc.* 60(2):271, 1972.
107. National Library of Medicine, "Orientation Handbook for New Members of the Board of Regents and the Biomedical Library Review Committee of the National Library Review Committee of the National Library of Medicine," 1980.
108. National Library of Medicine, *Programs and Services: Fiscal Year 1981* (Bethesda, Md.: National Library of Medicine, 1981).
109. National Library of Medicine, "Summary Minutes of the Ad Hoc Advisory Group Meeting on Information Needs of Health Professionals Concerned With Health Services and Health Care Delivery," July 16, 1976.
110. National Library of Medicine, "National Library of Medicine and the Issue of Cost Recovery for Services," unpublished, Dec. 16, 1981.
111. National Research Council, Division of Medical Sciences, *Communications Problems in Biomedical Research: Report of a Study* (Washington, D.C.: National Academy of Sciences, 1963).
112. Neufeld, M. L., Federation of Abstracting and Indexing Services, Philadelphia, "The Role of Secondary Services in the Transfer of Toxicology Information," unpublished, 1981.
113. "News Media and the Doctor-Patient Relationship," *N. Eng. J. Med.* 305(21):1278, 1981.
114. Oettinger, A. G., *Elements of Information Resources Policy: Library and Other Information Services* (Cambridge, Mass.: Harvard University, 1976).
115. Office of Technology Assessment, U.S. Congress, *Computer-Based National Information Systems: Technology and Public Policy Issues,* GPO stock No. 052-003-00852-1 (Washington, D.C.: U.S. Government Printing Office, 1981).
116. Office of Technology Assessment, U.S. Congress, "The National Library of Medicine," staff paper, March 1981.
117. Office of Technology Assessment, U.S. Congress, *Strategies for Medical Technology Assessment,* GPO stock No. 052-003-00887-4 (Washington, D.C.: U.S. Government Printing Office, September 1982).
118. Olson, P. E., "Mechanization of Library Procedures in the Medium-Sized Medical Library: A Study of the Interaction of Nonlibrarian Searchers With the MEDLINE Retrieval System," *Bull. Med. Libr. Assoc.* 63:35, 1975.
119. Pao, M., "A Quality Filtering System for Medical Literature," *J. Med. Educ.* 50:353, 1975.

120. President's Advisory Committee, "Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information," Jan. 10, 1963.
121. President's Commission on Heart Disease, Cancer, and Stroke, *A National Program to Conquer Heart Disease, Cancer, and Stroke* (Washington, D.C.: U.S. Government Printing Office, 1965).
122. President's Science Advisory Committee, *Improving the Availability of Scientific and Technical Information in the United States* (the Baker Report), W. O. Baker, Panel Chairman, 1958.
123. Relman, A. S., "Are Journals Really Quality Filters?" in *Coping With the Biomedical Literature Explosion: A Qualitative Approach,* K. Warren and W. Goffman (eds.) (New York: Rockefeller Foundation, December 1978).
124. *Research on Selective Information Systems* (New York: Rockefeller Foundation, 1980).
125. Resnikoff, H., Day, M., Farley, K., et al., *Report of the Working Group on Private Sector/Government Relationships for Scientific and Technical Information* (Washington, D.C.: National Science Foundation, June 1980).
126. "Resolution," Board of Regents of the National Library of Medicine, Oct. 30, 1981.
127. Rutkowski, R. M., "U.S. International Information Policy-Formation and Representation," presented before the Ninth Annual Telecommunications Policy Research Conference, Annapolis, Md., Apr. 27, 1981.
128. Sammons, J., "Medical Computing: A New National Agenda," an address to the Fifth Annual Symposium Computer Applications in Medical Care, Washington, D.C., Nov. 2, 1981.
129. Scura, G., and Davidoff, F., "Case-Related Use of the Medical Literature," *J.A.M.A.* 245:50, 1981.
130. Sewell, W., and Bevan, A., "Nonmediated Use of MEDLINE and TOXLINE by Pathologists and Pharmacists," *Bull. Med. Libr Assoc.* 64:382, 1976.
131. Sherman, J. F., "Report to the National Library of Medicine From the Task Force on Review of the Extramural Research Grant Program," unpublished, 1978.
132. Smith, K., "The Current On-Line User Policy of NLM," presented to the Board of Regents, National Library of Medicine, Washington, D.C., Oct. 29-30, 1981.
133. Solomon, R., and de Sola Pool, I., "Issue in International Access to Computerized Databases: The National Library of Medicine's MEDLARS," prepared under contract for the Office of Technology Assessment, U.S. Congress, December 1981.
134. Sternick, B., "MEDLINE Data Communications," *Med. Infor.* 3(3):185, 1978.
135. Stinson, E., and Mueller, D., "Survey of Health Professionals' Information Habits and Needs: Conducted Through Personal Interviews," *J.A.M.A.* 243(2):140, 1980.
136. Stowe, S. M., and Weiner, J. M., *The Application of Document Data Management Techniques to Quality Filtering of Scientific Literature,* proceedings of the 44th Annual Meeting of the American Society for Information Sciences (White Plains, N.Y.: Knowledge Industry Publications, 1981).
137. Summit, R., President, DIALOG Information Services, Inc., Palo Alto, Calif., personal communication, May 1982.
138. Surgeon General's Conference on Health Communications (Washington, D.C.: Public Health Service, February 1963).
139. Tagliacozzo, R., "The Consumers of New Information Technology: A Survey of the Utilization of MEDLINE," *J. Am. Soc. Info. Sci.* 26(5):294, 1975.
140. Tagliacozzo, R., "Estimating the Satisfaction of Information Users," *Bull. Med. Libr. Assoc.* 65:243, 1977.
141. Tagliacozzo, R., "MEDLARS: Its Utilization in Research, Education and Practice," *Methods Info. Med.* 12:170, 1973.
142. Task Force Report to the President's Special Assistant for Science and Technology, "Scientific and Technical Communications in the Government," J. H. Crawford, Panel Chairman, AD299-545, 1962.
143. U.S. Congress, Joint Committee on Printing, *Federal Government Printing and Publishing: Policy Issues,* report of the Ad Hoc Advisory Committee on Revision of Title 44 to the Joint Committee on Printing (Washington, D.C.: U.S. Government Printing Office, 1979).
144. U.S. Domestic Council Committee on the Right of Privacy, *National Information Policy,* report to the President of the United States (Washington, D.C.: U.S. Government Printing Office, 1976).
145. U.S. Federal Council for Science and Technology, Committee on Scientific and Technical Information, *Recommendations for National Document Handling Systems in Science and Technology,* NTIS No. PB 198 267 (Washington, D.C.: Federal Council for Science and Technology, Sept. 15, 1965).
146. U.S. House of Representatives, Committee on

Appropriations, hearings, Apr. 2, 1974 (Washington, D.C.: U.S. Government Printing Office, 1974).

147. U.S. House of Representatives, Committee on Government Operations, Subcommittee on Government Reorganization and Internal Operations, *Documentation, Indexing and Retrieval of Scientific Information* (Washington, D.C.: U.S. Government Printing Office, 1960).

148. U.S. House of Representatives, Committee on Interstate and Foreign Commerce, *Biomedical Research Extension Act of 1977* (Washington, D.C.: U.S. Government Printing Office, March 26, 1977).

149. U.S. House of Representatives, Committee on Interstate and Foreign Commerce, *Extension of Programs for Health Services Research and Statistics and Medical Libraries* (Washington, D.C.: U.S. Government Printing Office, 1976).

150. U.S. Senate, Committee on Appropriations, hearings, fiscal year 1976 (Washington, D.C.: U.S. Government Printing Office, 1976).

151. U.S. Senate, Committee on Appropriations, hearings, fiscal year 1977 (Washington, D.C.: U.S. Government Printing Office, 1977).

152. U.S. Senate, *Interagency Coordination of Information Hearings* (Washington, D.C.: U.S. Government Printing Office, 1963).

153. Van Camp, A., and Foreman, G., "BIOSIS PREVIEWS and MEDLARS—A Biomedical Team," *Online*, p. 24.

154. Verheijen-Voogd, C., and Mathijsen, A., "A Contribution to the Comparison of the Usefulness of the Data Bases of Excerpta Medica and MEDLARS in Biomedical Literature Retrieval," *Aslib Proceedings* 26(4):136, 1974.

155. Veterans Administration, Department of Medicine and Surgery, Washington, D.C., personal communication, Apr. 14, 1982.

156. Wanger, J., McDonald, D., and Berger, M., "Evaluation of the Online Search Process: A Final Report," contract No. N01-LM-8-4734, prepared for the National Library of Medicine, Jan. 31, 1980.

157. Warren, K., and Goffman, W., "Analysis of a Medical Literature: A Case Study," in *Coping With the Biomedical Literature Explosion: A Qualitative Approach*, K. Warren and W. Goffman (eds.) (New York: Rockefeller Foundation, 1978).

158. Warren, S., "National Library of Science Systems," memorandum to the President of the United States, 1967.

159. Weisberg, A., National Institute of Neurological and Communicative Disorders and Stroke, personal communication, May 1982.

160. Werner, G., "Use of On-Line Bibliographic Retrieval Services in Health Sciences Libraries in the United States and Canada," *Bull. Med. Libr. Assoc.* 67:1, 1979.

161. White House Conference on Library and Information Services, 1979, *Summary*, March 1980.

162. Willard, R., Vice President, Government Relations, Information Industry, testimony on House Bill No. 4758 before the Subcommittee on Government Information and Individual Rights, Feb. 25, 1982.

163. Williams, M., "An Analysis of On-Line Database Prices and a Rationale for Increasing the Price of MEDLINE," *Online Review* 6(1):7, 1982.

164. Williams, M., "Criteria for Evaluation and Selection of Bases and Data Base Service," *Spec. Libr.* 66(12):561, 1975.

165. Williams, M., "Information Science Research: The National Library of Medicine and the Public/Private Sectors," presented to the Board of Regents, National Library of Medicine, October 1981.

166. Williams, M., Chairman, National Library of Medicine, Board of Regents, presentation to the Board of Regents, National Library of Medicine, 1982.

167. Williams, M., and Lannom, L., *Databases Online in 1981* (Washington, D.C.: American Society for Information Science, 1981).

168. Zurkowski, P., President, Information Industry Association, Washington, D.C., personal communication, October 1981.

U. S. GOVERNMENT PRINTING OFFICE : 1982 98-764

MEDLARS AND HEALTH INFORMATION POLICY

ERRATA SHEET

Page	
11.	column 2, last para., line 3 ..... (36) instead of (142)
14.	Table 2, column 2 ..... Total records in data base instead of Total records at data base
15.	column 2, line 6 ..... more comprehensive instead of comprehensive
15.	column 2, 3rd para., line 4 ..... fuller description instead of full description
19.	Figure 2, 1st circle ..... commercial users 12% instead of commercial users 11%: 2nd circle ..... commercial users 11% instead of commercial users 12%
21.	column 1 3rd para., line 5 ..... trained instead of training
27.	column 1, 1st para., line 5 ..... (77) instead of (75)
27.	column 2, footnote, line 1 ..... (77) instead of (75)
28.	column 1, 2nd para., line 11 ..... American Public Health Association, "Report of the Ad Hoc Committee on Medical Care Terminology," unpublished, 1966-67 instead of (105)
41.	Table 9, column 1 ..... Major relevance to health care instead of Major relevance health care
51.	column 2, 2nd para., line 14 ..... (12,21,29,96,118) instead of (96,12,29,21)
53.	column 1, 3rd para., line 10 ..... data that are of interest instead of data that is not of interest
54.	column 2, 3rd para. line 5 ..... (40) instead of (49)
55.	column 2, 2nd para., line 8 ..... (168) instead of (66)
56.	column 2, line 2 ..... add (137) after service
57.	column 2, 3rd para., line 10 ..... add (55) after practices
59.	column 1, 1st para., line 15 ..... (168) instead of (66)
61.	column 2, 2nd para., line 2 ..... use fee rate instead of use fee ratio
65.	column 2, line 3 ..... as will be discussed further instead of as will be discussed
66.	column 1 3rd para., line 6 and 7 ..... there are no simple answers instead of there is no simple answer
79.	column 1 , 2nd para., line 10 ..... the other private instead of the other
92.	column 2, 2nd para., line 7 ..... U.S. Congress House Report 1026 accompanying The Medical Library Assistance Act of 1965 instead of (78)
98.	column 2, 4th para., line 14 ..... (91) instead of (94)
102.	column 2, 5th para., line 4 ..... add (125) after technology
103.	column 2, 2nd para., line 2 ..... add (96) after NCLIS
104.	column 1 4th para., line 9 ..... for complete 1981 data instead of for complete data
105.	column 2, line 4 ..... (129) instead of (14)
106.	column 1, 2nd para., line 4 ..... (63) instead of (11)
107.	column 1, 4th para. line 4 ..... (25) instead of (125)
110.	column 2, line 4 ..... (167) instead of (166)
114.	column 1 line 6 ..... (96) instead of (97)
139.	column 2, line 25 ..... Jonathan Ruby instead of Johnathan Ruby
139.	column 2, line 29 ..... Charles W. Sargent instead of Charles W. Sergeant
139.	column 2, line 36 ..... Thomas Jefferson University instead of University of Pennsylvania Medical School

BEST AVAILABLE COPY